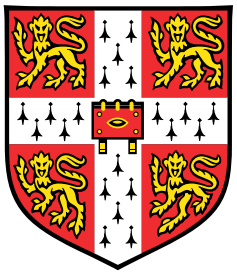
A modelling perspective on precipitation in the Indus River Basin: from synoptic to Holocene variability



Jean-Philippe Pierre André Baudouin

Michael Herzog

Cameron A. Petrie

Department of Geography
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

St Edmund's College

September 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. This dissertation has not been submitted in whole or in part for a degree or qualification at any other university. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the Geography Degree Committee.

Jean-Philippe Pierre André Baudouin
September 2020

Acknowledgements

First and foremost, I shall thank warmly Michael Herzog and Cameron Petrie, my two supervisors, without whom this project would have never started. Michael never lost the objective of my PhD research; I am particularly grateful for his help to put my thoughts and findings into the words you are about to read. Cameron's support and listening has been essential in moments of doubt. I have also found in his passion for the Indus Civilisation a fascinating fount of knowledge.

The TwoRains group has been an exciting place for multidisciplinary exchange, as I dreamt of before joining it, despite the communication difficulty between researchers from different fields. Alena Giesche, Andreas Angourakis, Joanna Walker, Cemre Ustunkaya, Nathan Wright, Adam Green, Hector Orengo and many others have brought ideas, discussions, cakes and support on countless Friday morning meetings.

In the department of Geography, the "informal-physical-geographer-third-year-PhD-student-group" with Oliver Taherzadeh, Annemarie Eckes-Shephard and Hyesop Shin offered a space for discussing our work and our research management in depth. A special thanks also goes to Peadar Brehony for his insight on human resilience and adaptation, as well as to Eimear Dunne, Celine Vidal, Thomas Aubry, Makoto Takahashi and Sibylla Warrington.

Finishing my PhD would not have been possible without the help of Kira Rehfeld, who offered me a post-doctoral position at the University of Heildelberg within the Stacy group while making sure I had enough time to complete my PhD research. Her insights on past climate variability were also crucial for the last part of my research. I also thank for their help the Stacy members, and particularly Janica Bühler, Beatrice Ellerhoff, Nils Weitzel, and Elisa Ziegler.

I am also grateful to Pascale Braconnot, Andrew Turner, and Louis-Philippe Caron to have invited me at their respective institute so I can present and discuss my work. It is thanks to them that I was able to meet with Kieran Hunt and Martin Menegoz that shared with me their insights on WD research and precipitation dataset uncertainty respectively. In that respect, PAGES offers an unrivalled platform for scientific exchange on past climate, and I would like to express my gratitude to its organisers for maintaining the links within such a wide community.

I would also like to thank the various groups which produced observational datasets and climate simulations and made them freely available online. None of this thesis would have been possible without all their prior hard work.

Last but not least, I am indebted to the European Research Council which funded my PhD stipend, as well as to St Edmund's college for its travel grants.

Abstract

The Indus River civilisation was the first urban society in South Asia. Its demise, ca. 4000 years ago, might have been caused by environmental factors, such as recurrent drought conditions. Yet, palaeoclimate archives present a fragmented picture of the climate at that time. This thesis explores the potential value of global climate models to further interpret the archaeological context. Precipitation variability is investigated at various timescales due to their different impact on human societies and their importance in the climate system: the synoptic scale, the seasonal cycle, the inter-annual to multi-decadal variability, and the multi-millennial Holocene trends. The precipitation from several climate model simulations is evaluated at each of these timescales.

Indus River Basin precipitation is first explored in observational datasets. This study highlights the quality of ERA5 reanalysis, which is used as a reference in the subsequent chapters. Statistical tools show that more than 80% of the precipitation in the Upper Indus Basin is related to cross-barrier moisture transport along the Himalayan foothills. The climate models analysed (IPSL-CM6-A, MRI-ESM2-0 and GISS-E2-1-G) generally reproduce this process well, but the seasonality of cross-barrier moisture transport is biased, resulting in precipitation biases: the main wet season, the summer monsoon, is shorter and significantly dryer, while the second wet season, in winter, is longer and more active.

The link between winter precipitation, cross-barrier moisture transport and Western Disturbances is further explored, first in ERA5 and then compared to IPSL-CM6-A model output. Sub-daily resolution is needed to determine the origin of the positive precipitation bias in winter. This bias is related to differences in the atmospheric circulation associated with Western Disturbances. The stronger Subtropical Westerly Jet is also key to understand the precipitation overestimation. Centennial to millennial-scale precipitation variability is more difficult to evaluate due to the short length of observations and the paucity of climate records, but the results suggest that inter-annual variability in the IPSL climate model family is overly dominated by atmosphere-only processes, with a potentially large impact of the precipitation response to external forcings.

In addition to biases in the representation of mean precipitation and synoptic to inter-annual variability, climate models are also limited by the representation of other internal

processes such as ocean circulation, and vegetation-dust aerosols, as well as by the uncertainty in some external forcings such as volcanic eruptions and the solar activity. Hence, at the present state, past-climate simulations do not provide the nuanced information of precipitation changes and variability that is needed to understand impact of precipitation variability on archaeological contexts, and will not do so until significant breakthroughs are achieved. Nevertheless, climate models remain a powerful tool for climatologists to investigate large-scale processes that can eventually better characterise climate variability.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation and scope

Water is a fundamental nexus in human societies. Throughout human history, failures in anticipating water shortages or flooding events have threatened human activities and most importantly food production. Human populations have adapted their behaviour within specific environments to increase their resilience in the face of this climate variability. However, the efficiency and even the success of these responses remain an open question, especially in case of extreme climate events. A failure to mitigate such extremes can have dramatic societal consequences.

TwoRains[1] is a project investigating the specific resilience and sustainability of the Indus Civilisation, the first South Asian complex society (cf. Figures 1.5 and 1.6). The TwoRains team includes a wide range of archaeological expertise, from archaeobotany and geoarchaeology to ceramic analysis, as well as climate competence. The aim is to build a multi-disciplinary understanding of Indus society and its relation to its environment, in order to better understand the establishment and the mysterious demise or decline of the civilisation about 4000 years ago, for which both natural and human factors have been suggested as credible causes (see Petrie, 2017). In particular, the project investigates the agricultural practices and the consumption habits of the Indus population of that period, in relation with the underlying settlement distribution. It also aims to reconstruct the past environmental and climate conditions in the Indus River Basin with a specific focus on rainfall patterns which this thesis addresses.

Past climate information is generally derived from climate archives, but these present disadvantages. The age uncertainty of these datasets often reaches 100 years for the period

[1] http://www.arch.cam.ac.uk/research/projects/two-rains

considered, which makes comparison of multi-decadal events between records difficult. Climate records have a complex relation to climate variables (e.g. precipitation or temperature) and can be influenced by other environmental factors. They are also sparse in space and provide information on the local climate conditions often far away from the archaeological sites. They also rarely provide information on the seasonality. By contrast, climate model outputs directly provide information on the climate variables, including their spatial pattern and seasonality. While based on our physical understanding of climate processes, climate model often produced biased output, and also rely on an accurate representation of the boundary conditions (e.g. land surface, solar radiation). This thesis focuses on the potential added-value of model data being used in conjunction with climate records, and it revolves around the evaluation of climate model outputs, and relevant methods to do this in the context of TwoRains.

1.2 The geography of the Indus River Basin

1.2.1 Relief

The Indus River Basin has a very contrasted relief (Figure 1.1). The basin is bounded from the north-east to the west by high mountain ranges, including the Himalayas, Karakoram, Hindu Kush, and Sulaiman ranges. The northern part of the basin actually includes seven of the world's 20 highest mountains such as the K2, the Nanga Parbat or the Nanda Devi. The basin is separated from peninsular India by the lower Aravalli range. This range does not extend up to the Himalayas, making no obvious boundary between the Indus and the Ganges watershed, an area that is on the contrary very flat. The same hill-less low altitude plain dominates most of the inner and southern Indus Basin. The mountain ranges to the north are still tectonically active, with uplift rates between 2 to 12 mm per year (Burbank 1996), but these rates are not high enough to significantly change elevations across the Holocene.

1.2.2 Climate

This contrasted relief leads to a large variety of climate type across the basin (cf. Köppen climate classification; Petrie et al., 2017: Figures 2-A and B]. Altitude dependant temperature patterns overlap with mostly slope dependant precipitation patterns (Figures 1.2 and 1.3). The Thar Desert dominates much of the southern flat plain, with temperatures reaching as high as 50°C each year (Kothawale et al., 2010) and precipitation below 300 mm per year on average (Figures 1.2-A). Wetter, semi-arid climate develops towards the Aravalli range and the Himalayas. By contrast, heavy precipitation falls on the southern slope of the Karakoram

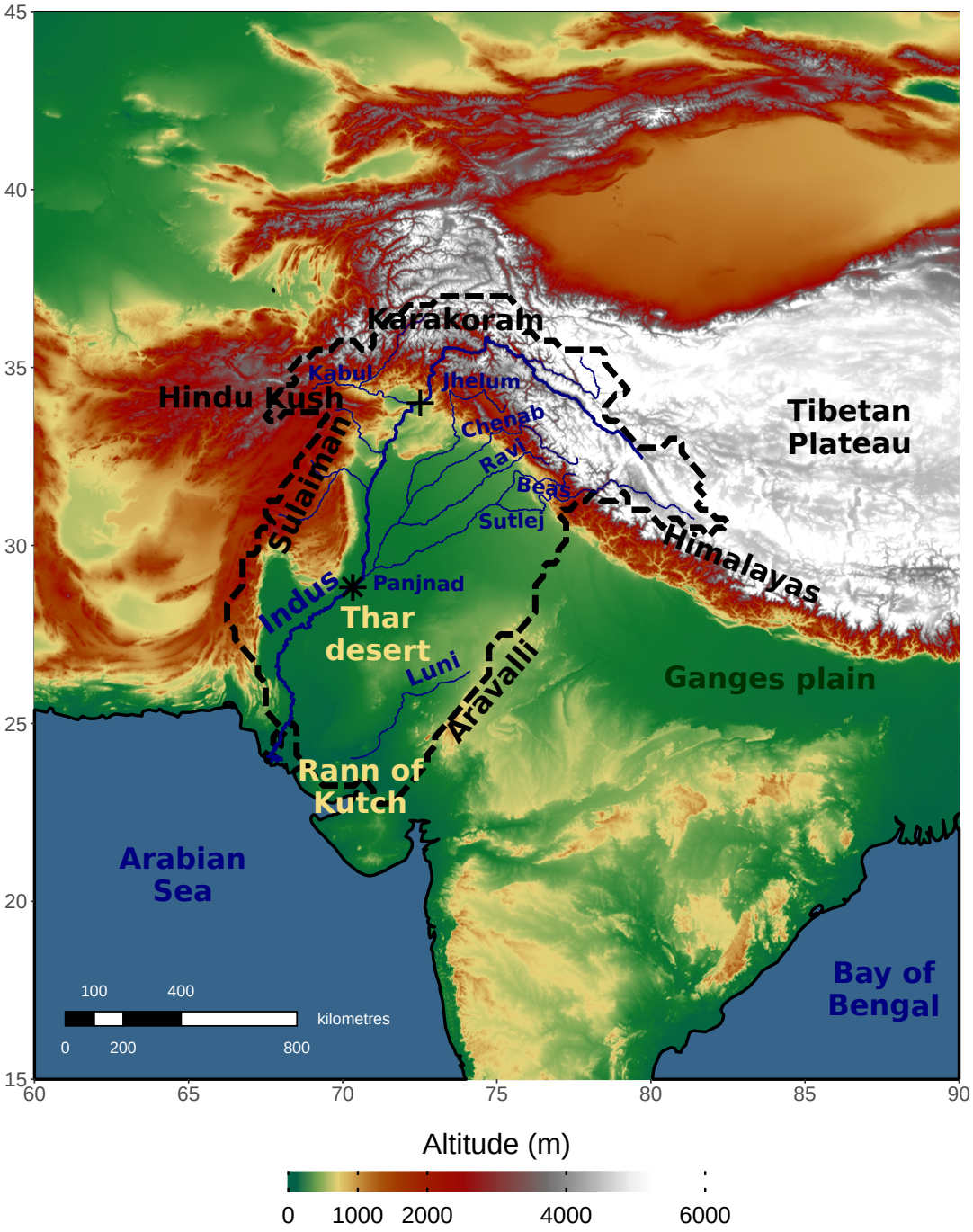


Fig. 1.1 Relief and main geographical features in the Indus River Basin and its surroundings. The star marks the location of the Sukkur dam and the plus that of the Tarbela dam.

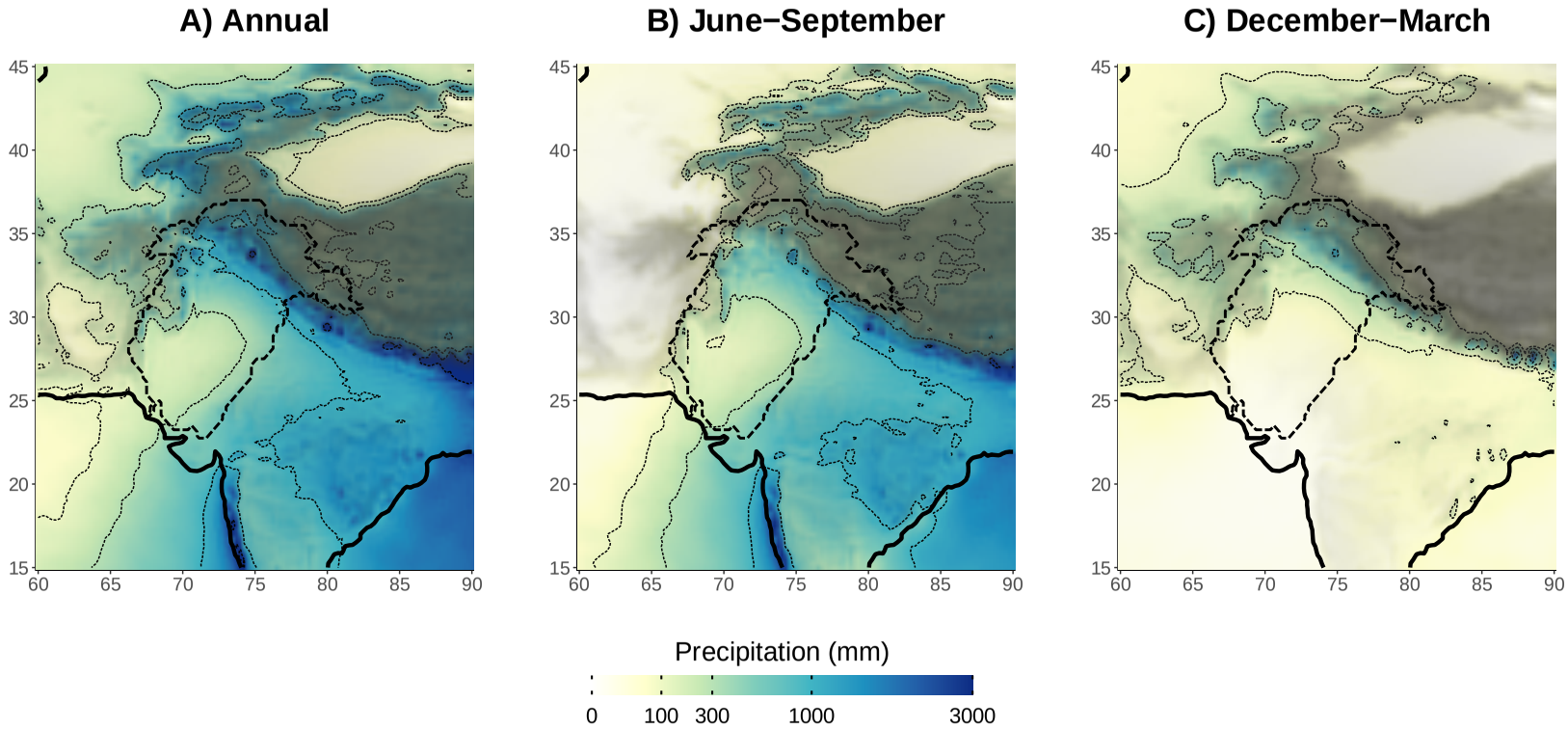


Fig. 1.2 Map of the mean precipitation deriving from ERA5 reanalysis (1979-2018) as an average over the whole year (A), summer (B) and winter (C). The dotted contour lines correspond to the 100, 300, 1000, and 3000 mm isohyets. The dashed contour is the Indus River Basin, and the grey shading correspond to the relief as modelled in ERA5.

and the Himalayas foothills (Siwalik hills, Punjab) leading to a gradation of the climate from humid subtropical to humid continental, alpine tundra and glaciers or ice cap depending on the altitude. The deep valleys further to the North of these ranges are in the rain shadow and receive little precipitation establishing a cold arid or cold semi-arid climate.

A complex seasonality on both temperature and precipitation is evident (Figures 1.4). Coldest temperatures occur in January with frequent frost and snowfall at 1500 m altitude and above (Gurung et al., 2017). Frost can even happen at low elevation in the plain and endanger crops (De et al., 2005; Yadav et al., 2012). The hottest temperatures occur in June over the plain with sometimes a second peak in September-October. The less warm temperature in July and August are related to an increase in precipitation. Summer is indeed the wettest season basin wide with precipitation extending from India through the onset of the South Asian monsoon (Figure 1.2-B). A second precipitation peak also occurs in winter but is limited to the mountains and the foothills to the north and the north-west (Figure 1.2-C).

Other climatic phenomena of note include tropical cyclones close to the Arabian Sea coast in May-June and October-December, dense fog in winter, and thunderstorms, associated with heavy winds, tornadoes and large hail over the plain, and flash floods in the mountains, particularly between February and August (De et al., 2005).

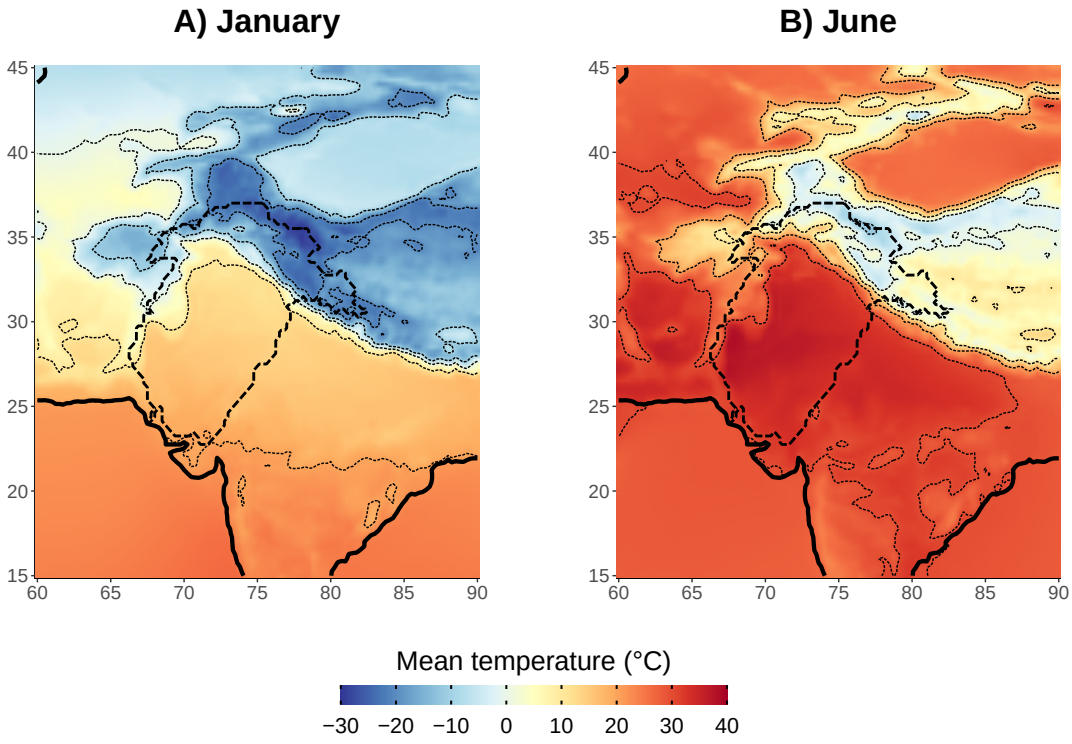


Fig. 1.3 As Figure 1.2, but with mean surface temperature for January (A) and June (B)

1.2.3 Hydrology

The Indus Basin can be split in four hydrological areas: the upper Indus, the western Indus, the Punjab region, and the lower Indus.

The upper Indus includes the part of the river and its tributaries from its source, at 4200 m of altitude on the Tibetan plateau, in China, up to the Tarbela dam, before the confluence with the Kabul River, at a much lower altitude of 400 m (cf. the "+" on Figure 1.1). The Indus River forms there a deep and narrow valley between the Karakoram and the Himalayas, flowing westwards into Indian Laddakh, before turning south and emerging in the plain. The valleys themselves have a dry climate, but they gather water from the mountains, where wetter conditions dominate (Immerzeel et al., 2009, 2015). The mean discharge is about 2300 $m^3 s^{-1}$ at the Tarbela dam, but up to 80% of the run-off occurs between June and September (Yu et al., 2013). Most of the run-off derives from snow and glacier melt and thus occur during the warmest months (Archer and Fowler, 2004).

The Indus River continues then to the South, while gathering its right-bank tributaries from the Hindu Kush and the Sulaiman ranges and forms the western Indus region mentioned

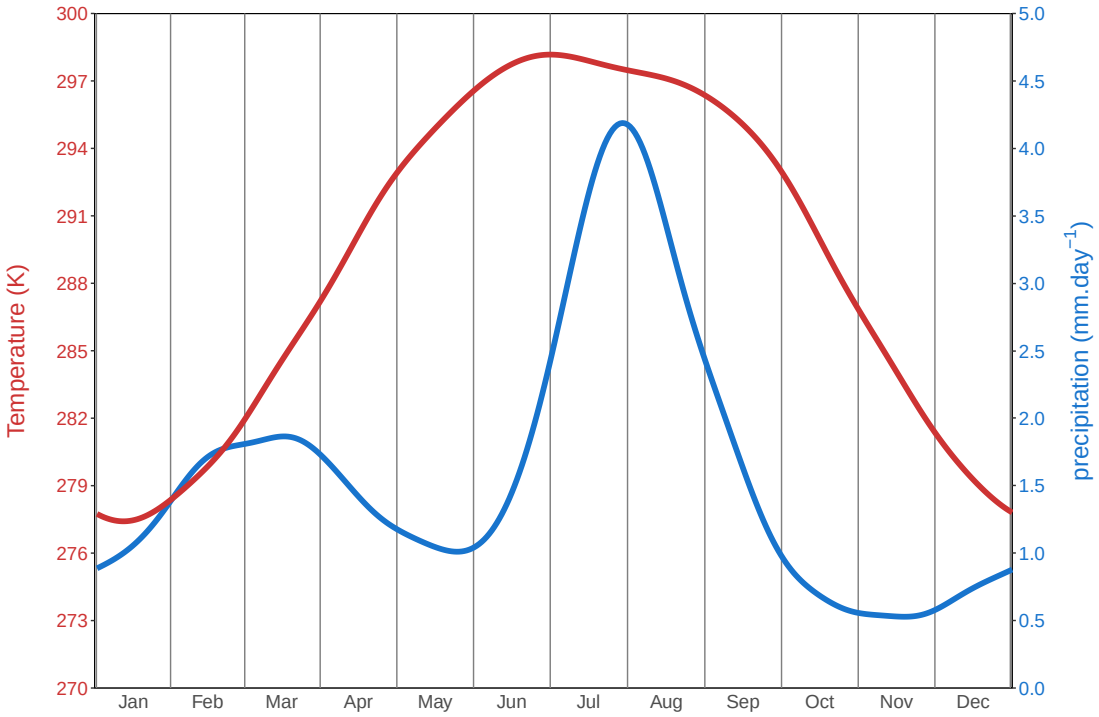


Fig. 1.4 Mean seasonality of precipitation and temperature as an average over the Indus River Basin (ERA5, 1979-2018)

previously. The discharge seasonality follows a similar pattern, but with stronger discharge in spring due to the higher contribution from winter snow melt than from glaciers (Casale et al., 2019; Hashmi et al., 2019).

On the eastern side of the Indus River, all the five main left-bank tributaries converge into one river (The Panjnad) before reaching the Indus (Figure 1.1). They account for between one third (Karim and Veizer, 2002) and up to a half (Murakami, 1995) of the total Indus discharge. They flow in the so-called region of the five rivers, Punjab, which sees heavy precipitation during the summer season, but also during winter. In fact, all rivers of that region have a mostly pluvial regime, apart from the Sutlej for which half of the discharge comes from snow and ice melting before the confluence with the Beas River, as it flows deeply in the inner Himalayas (Singh and Jain, 2002). The Panjnad reaches the Indus just upstream the Sukkur dam (cf. the star on Figure 1.1). The downstream part of the Indus River, from that dam to the mouth at the Arabian Sea, forms the lower Indus. No perennial tributary reaches the Indus there and this is the driest part of the basin. Consequently, the discharge of the Indus River is maximum after the confluence with the Panjnad River, with a mean around $6500\,\mathrm{m^3\,s^{-1}}$ (Murakami, 1995) and large seasonal amplitude. For example, during the 2010 record breaking flood, the runoff peaked at more than $32\,000\,\mathrm{m^3\,s^{-1}}$, above the dam capacity (Khan et al., 2011). Furthermore, the extensive irrigation network developed in Pakistani and Indian Punjab over the last decades has led to a severe decrease of the discharge at the mouth of the Indus River (Karim and Veizer, 2002; Wescoat Jr. et al., 2018).

A fifth zone, east of the Indus and Sutlej Rivers, is included in the study area (Figure 1.1). This zone encompasses large part of the Thar Desert, between the Aravalli range and the Indus and Sutlej Rivers. Most of the sparse rain in this region does not generally reach the Indus River. It rather evaporates in the desert or reaches the Rann of Kutch, a flat salt marsh with a complex connection with the Arabian Sea and the Indus mouth (Syvitski et al., 2013) through the seasonal Luni River (southeastern river in Figure 1.1). Thus, this zone is not strictly part of the Indus watershed, but is included in the study area since the Aravalli range provides a clearer boundary. In addition, one of the Indus cities (Dholavira, cf. Figure 1.5, Section 1.3.1) is present in the middle of the Rann of Kutch.

Seasonal variations of river levels in the flat alluvial Indus plain lead to an unsteady hydrological network. Rivers change their course, sometimes abruptly, as illustrated by the avulsion of Kosi River that occurred in August 2008 in Northern India (Sinha, 2009). Many past paleo-channels have been detected by the TwoRains project (Orengo and Petrie, 2017, 2018). Those gradual or swift changes in river paths are likely caused by neotectonic activity or extreme rain outbreak (Giosan et al., 2012; Paliwal, 2013; Syvitski et al., 2013). Hence,

the landscape could have changed because of this variable hydrological network, and could have influenced the history of the Indus Civilisation (Possehl, 1999).

1.2.4 Landscape and vegetation

Landscape and vegetation are adapted to the diversity of terrain and hydro-climatic conditions. The glaciers and bare ground at the highest altitude progressively evolve into a tall temperate forest at lower altitude on the wet slope of the Himalayas (Himalayan Moist Temperate Forests; Ashutosh, 2019). The foothills and the plain of Punjab are mostly irrigated crop field today, but a gradation of natural vegetation still exists, from tropical deciduous to thorn forest (Ashutosh, 2019) and eventually dry grassland (ul Islam and Rahmani, 2011) as precipitation decreases to the south. The bare ground of the Thar Desert dominates most of the central and southern plain away from rivers and irrigation canals with numerous old and active dune formations (Czerniawska and Chlachula, 2017; Singhvi and Kar, 2004). Since the early to mid-Holocene, the desert has extended through the advancement of dune fields and the desiccation of lakes (Bryson and Swain, 1981; Dixit et al., 2018; Enzel et al., 1999), suggesting a change in the hydro-climatic conditions, further discussed in Section 1.4.1. Climate and human impacts are also evident on the landscape near the foothills (Demske et al., 2016).

1.3 The Indus Civilisation

1.3.1 Overview

Figures 1.5 and 1.6 set the spatial and temporal context of the urban phase of the Indus civilisation. The earliest settlements in the Indus region originally appeared on alluvial fans towards the western part of the Indus Basin (Balouchistan on Figure 1.5) around 9000 to 7000 years ago (Figure 1.6; Petrie and Thomas, 2012). They were occupied by farmers who used domesticated grains (wheat and barley) and livestock (bovine, caprine and ovine), many of which were imported from the Fertile Crescent and the Iranian Plateau (Figure 1.6; C, 1996; Petrie, 2015). There is also evidence of local domestication, such as the zebu (Bos Indicus, Chen et al., 2009). Subsequently, settlements started to appear on the Indus River plain, from Punjab to Sindh (Figure 1.5), as the population grew (Regionalisation era; 1.6; Possehl, 1998, 1999) and the exchange of raw materials and finished products expanded (Law, 2011). This expansion of settled population was also accompanied by an enhanced sophistication in craft production through the use of new materials and finer and more complex decoration methods (Vidale, 2000). Ultimately, urban planning was developed as proven by the construction of

fortification walls, drains and wells (Kenoyer, 1997, 2008; Wright, 2010). The pre-urban phase (ca. 5300-4600 BP, Figure 1.6) saw the development of bigger towns (e.g. Kot Diji, Harappa), the extension of the exchange network to neighbouring areas (to import lapis lazuli for instance, Law 2014), diversification of food production and crafts, or creation of wealth such as jewellery (Kenoyer, 1997). However, it was only after ca. 4600 BP that some settlements enlarged to city size (over 50 ha, namely Mohenjo-Daro, Harappa, Rakhigarhi, Dholavira and probably Ganweriwala; Figure 1.5; Kenoyer, 2008; Petrie, 2013). During this urban phase (ca. 4600-3900 BP), the cultural practices became more uniform (Kenoyer,

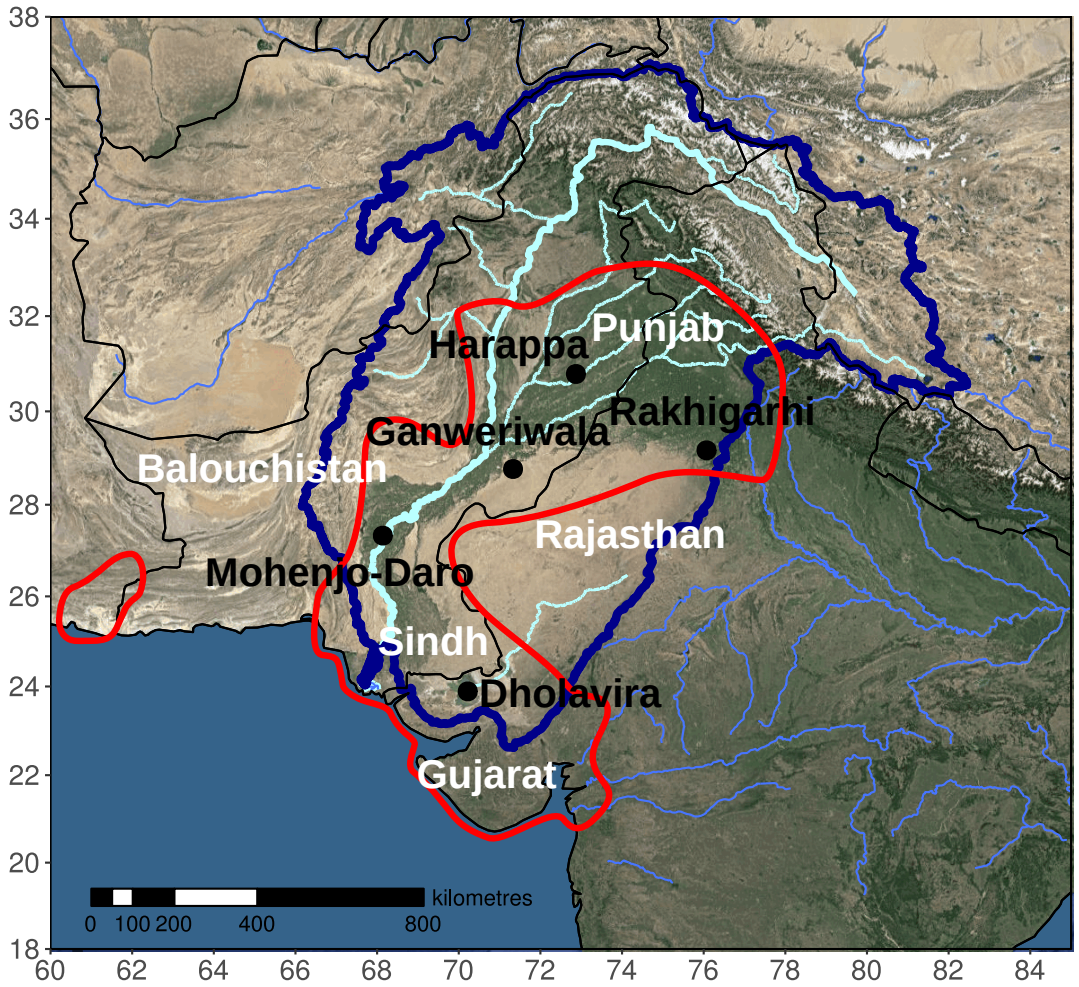


Fig. 1.5 Extension of the Indus civilisation during the urban period (red contour) with the five main settlements (black dots). The background shows modern coastlines, rivers and satellite images. The blue contour is the modern extension of the Indus River Basin. Historic region names are also indicated in white.

1997; Wright, 2010) with the widespread use of the emblematic Indus seals and weight measurement system. Trade and exchange along the Indus River and between the cities was intense, and contacts were made with populations on the Arabian Peninsula, in Mesopotamia, Central Asia, and other part of India (Law 2006). In spite of the evidence for an overtly culturally-integrated Indus Civilisation, there is no conclusive proof of the constitution of a centralised state (Kenoyer, 2008; Possehl, 1998), as for its contemporary urban civilisations (Old Egyptian Kingdom, Akkadian Empire in Mesopotamia, and latter Chinese Shang dynasty, Figure 1.6). Despite the standardisation, a degree of cultural variability still existed across the Indus territory during the urban phase (Petrie et al., 2018, 2017; Possehl, 2002), which may suggest a more decentralized, rural-based society organised around only a few urban centres compared to other early urban societies (Petrie, 2013).

1.3.2 Relation to the climate

Up to the modern period, the Indus population have adapted their agricultural practices to the climate conditions and particularly the occurrence of two spatially overlapping wet seasons (Figures 1.2-A,B, and 1.7). Unlike in any other early farming areas, both winter crops (or "rabi", mainly wheat and barley) and summer crops (or "kharif", mainly millet, beans, and rice) were cultivated allowing for two harvest seasons. The diversity of agricultural practices needed to exploit winter and summer cereals, pulses, and fruits have therefore been described as "multi-cropping" (Petrie and Bates, 2017; Weber, 2003).

The risk of crop failure is dependent on the local scale intra-seasonal to inter-annual variability of precipitation and water availability. The summer is a critical season, when both floods and droughts can occur and endanger crop yields. Hence, multi-cropping might have helped to buffer risks for food production in a settlement, especially in north-west India where summer and winter precipitation distribution overlap (Figure 1.7) and no specific

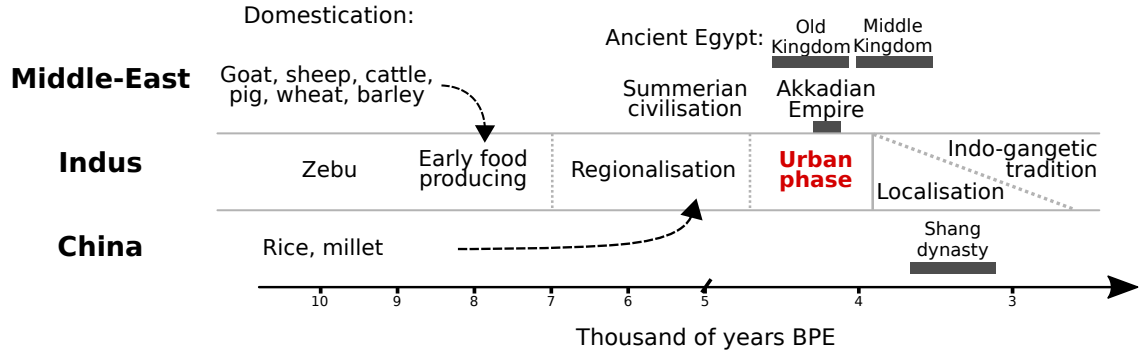


Fig. 1.6 Chronology of early human societies, with a focus on the Indus plain.

cultivating types predominated (Petrie and Bates, 2017). Some water management is also evident (e.g. reservoirs; Walker in prep.) that could have helped mitigating at least the seasonal drought in spring. Furthermore, the urbanisation process may have resulted in the regional mutualisation of resources that mitigated the impact of local crop failure in one or a few settlements (Petrie et al., 2018; Petrie and Bates, 2017). The extension and importance of these systems for mitigating climate extremes remain largely unknown and are one of the core questions being explored by TwoRains. This thesis contributes to this discussion by providing information on the intra-seasonal to inter-annual variability of the precipitation.

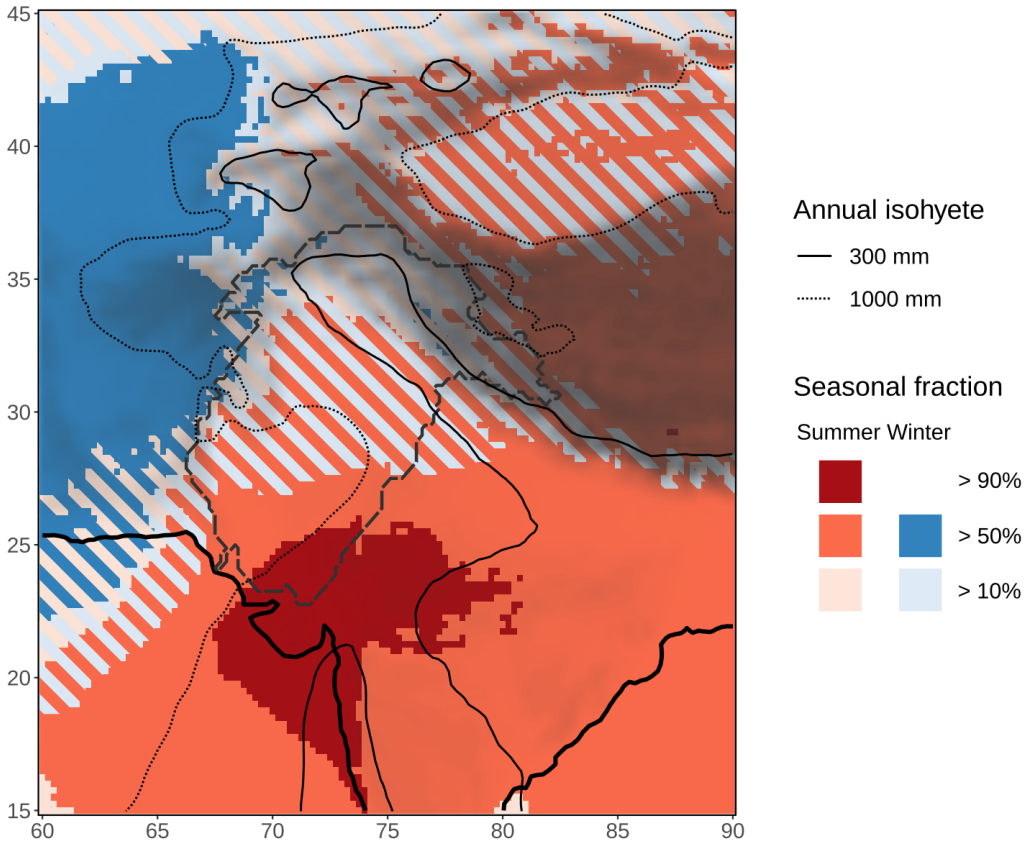


Fig. 1.7 Map of the overlap of winter and summer precipitation in the Indus River Basin. The colour shading indicates the fraction of annual precipitation falling in summer (June-September) and in winter (December-March). Hatching indicates areas where at least 10% of the precipitation occurs both in winter and in summer. Based on ERA5 (1979-2018)

1.3.3 Demise

Archaeological evidence shows that a major shift in settlement location had occurred by 3900 BP and put an end to the urban phase of the Indus Civilisation. All the cities declined and were progressively abandoned, while many other smaller settlements in the Indus River Basin were displaced. The settlement density remained high only in Gujarat and eastern Punjab, although archaeological surveys have not been comprehensive (Petrie et al., 2017). This shift was accompanied by a drop in the range and intensity of trade (Law, 2006), while elaborated crafts were no longer produced (Possehl, 1997a,b). Both natural and human factors have been conjectured as credible causes of this process, either described as a "collapse", a "decline", or a "transformation" (Possehl, 1997a,b; Wheeler, 1968), though there is no agreement about which factor is most significant (Possehl, 1997b). Petrie (2017, 2019) and others have developed the idea of "transformation" as to oppose to the concept of "crisis to collapse" that is more often developed in the literature. The main difference between the two is the time-space scale of change that is important to understand the causes.

Climate change has potentially dramatic impacts on agricultural production and food security through the variation of water supply (Al et al., 2008). However, there is no consensus whether a change in precipitation has affected the population, some believing that there is no conclusive evidence (Kenoyer, 2008; Possehl, 1997a), while others present this as the main factor (Clift and Plumb, 2008; Giosan et al., 2012; Staubwasser and Weiss, 2006). However, there is growing evidence for a shift or increased variability in climate around 4100 BP (4.2 event, cf. Section 1.4.1), although no direct link has yet been established with archaeological observations (Petrie, 2017).

Migrations and changes in the settlement distribution are an efficient but dramatic response to lasting changes in the social or natural environment. However, such changes imply a loss of infrastructures such as those related to water management, cities, trading networks, and the cultural practices related to them. While these infrastructures may mitigate the intra-seasonal to inter-annual variability of the climate, their rigidity against migration make them more vulnerable to shifts or multidecadal events or variability in climate. Hence, another core question of TwoRains relates to the resilience and sustainability of the spatial and social organisation of the Indus society towards shifts (e.g. river avulsion, drying trend) and low frequency variability (e.g. 4.2 event, cf. Section 1.4.1) in the environment (Green and Bates et al., in press; Petrie et al., 2017). Consequently, this thesis also addresses the question of the variability of precipitation up to the millennium scale. The range of mobility indicated by the settlement distribution change also suggests a spatial scale for the study of about 100 km.

1.4 Climate variability during the Holocene

1.4.1 Inferred by precipitation proxies

Past climate changes can only be inferred indirectly through climate archives that record environmental changes. Proxy studies include the chemical analysis of marine or lake sediments, speleothems, ice cores, corals, but also dendrochronology, palynology and geology. Detection of isotopic oxygen ratios in speleothem is the most promising method to infer precipitation thanks to high temporal resolution (up to a few years) in otherwise poorly documented tropical and subtropical areas (Lachniet, 2009). Pollens studies are also key as they allow for the reconstruction of land cover variability. Uncertainties are higher for climate records than for direct observations, not only in the magnitude of change recorded, but also in its dating. Hence, analyses are generally limited to qualitative changes. A short review of Holocene climate records in relation to the Indus River Basin will now be presented.

Despite the lack of precipitation proxies close to archaeological sites in the alluvial plains of the Indus River and its tributaries as noted by Petrie and Weeks (2019), a growing number of studies have been published using records from the edges of the Indus River Basin, particularly over the last decade. Giesche (2020) noted that these studies are clustered in four areas, where environmental and geological conditions are favourable to retrieve climate records. The first climate analyses were based on records from the Thar Desert playas and lakes with more recent studies further expanding on the desert margins (Achyuthan et al., 2007; Bryson and Swain, 1981; Deotare et al., 2004; Dixit et al., 2018, 2014a,b; Enzel et al., 1999; Prasad et al., 1997, 2014b, 2007; Raj et al., 2015; Roy et al., 2009; Singh et al., 1972, 1974, 1990; Sinha et al., 2006; Swain et al., 1983; Wasson et al., 1984). Then, marine cores were retrieved from the Indus fan, in the north-east of the Arabian Sea (Azharuddin et al., 2017; Burdanowitz et al., 2019; Doose-Rolinski et al., 2001; Giesche et al., 2019; Giosan et al., 2018; Kahkashan et al., 2020; Lückge et al., 2001; Sirocko et al., 1993; Staubwasser et al., 2002, 2003; von Rad et al., 1999). Records have also been collected in high-altitude Himalayan lakes and peats sediments (Bhushan et al., 2018; Chakraborty et al., 2006b; Demske et al., 2016, 2009; Dutt et al., 2018; Leipe et al., 2014; Mishra et al., 2015; Nakamura et al., 2016; Phadtare, 2000; Prasad et al., 2016; Rawat et al., 2015; Srivastava et al., 2017; Trivedi et al., 2013; Wünnemann et al., 2010). Finally, the most recent efforts have concerned speleothems records, generally close to the Himalayas (Berkelhammer et al., 2013; Denniston et al., 2000; Joshi et al., 2017; Kathayat et al., 2018, 2017; Kotlia et al., 2018, 2015; Sinha et al., 2015).

Most of these studies agree that wet conditions dominated the early to mid-Holocene, following the deglaciation and the rise in temperature (Kaushal et al., 2018; Misra et al.,

2019; Roy and Singhvi, 2016). This wet optimum was followed by a long-term drying trend between 6 ka and 2 ka BP, from which modern conditions were established. These trends have mostly been attributed to fluctuations in the South Asian summer monsoon in both intensity and westward extension. One exception is the pollen studies of the Thar Desert and its margin, which have suggested that winter precipitation was as important for explaining the mid-Holocene wet optimum (Bryson and Swain, 1981; Prasad and Enzel, 2006; Roy and Singhvi, 2016). This hypothesis is surprising as this region is the least affected by winter precipitation nowadays (Figures 1.2-C and 1.7). These authors might have been referring to post-monsoon precipitation.

The lack of winter specific precipitation proxies in the Indus River Basin is an acute issue for archaeological studies, which make do with proxies from elsewhere despite the potential differences with the Indus Basin conditions (Petrie and Weeks, 2019). Interestingly, proxies from both Central Asia (Aizen et al., 2016; Chen et al., 2008) and the Middle-East (Andrews et al., 2020; Cheng et al., 2015), which are affected by extra-tropical precipitation, show the establishment of wetter conditions in the early to mid-Holocene, followed by a drying trend, similar to the summer monsoon proxies.

The drying trend after the Holocene optimum was not gradual everywhere and some abrupt changes occurred at the multi-decadal to centennial scale. The 8.2 ka BP event was primary characterised by cooler temperatures in the North Atlantic for about one or two centuries (Johnsen et al., 2001) , due to changes in the ocean circulation (Matero et al., 2017). This event is recorded in the Asian summer monsoon domain as a dry spell, including in Arabia (Gupta et al., 2003; Parker et al., 2006), in the Thar desert (Dixit et al., 2014b), and up to China (Hong et al., 2005). In the Thar Desert, the drying of lakes was often abrupt (Roy and Singhvi, 2016), but the most notable event occurred around 4.2-4 ka BP (Dixit et al., 2018, 2014a). This so called 4.2 event has been discussed for other Indus records, and particularly the Indus fan (Giesche et al., 2019; Staubwasser et al., 2003). A drought is also recorded in other parts of the South Asian summer monsoon domain, such as south India (Sandeep et al., 2017), central India (Prasad et al., 2014a) and northeast India (Kathayat et al., 2018), but interestingly not from the Tibetan Plateau. In addition, signs of droughts are also evident from the west of the Indus, such as in the Persian Gulf (Cullen et al., 2000), the Red Sea (Arz et al., 2006) or the Iranian plateau (Carolin et al., 2019) where winter precipitation dominates. The 4.2 event might have been global, as suggested by a climate archive from the Kilimanjaro (Thompson et al., 2002), Namibia (Railsback et al., 2018), the Mediterranean Sea (Kaniewski et al., 2018), and Iceland (Geirsdóttir et al., 2019). However, the exact timing, duration, seasonality and extension of the 4.2 event drought in the winter westerlies and summer monsoon domains is still debated (Giesche, 2020; Kathayat et al.,

2018). Some proxies suggest the occurrence of several droughts around that time (Giesche et al., 2019; Kathayat et al., 2018), while others suggest that it was embedded within a period of high multi-decadal variability with regional specificity (Kathayat et al., 2017; Misra et al., 2019; Srivastava et al., 2017).

1.4.2 Drivers of the Holocene variability

1.4.2.1 External forcings

Paleo-climatologists investigate the causes of past climate variability using both climate records and climate models. These causes are of two types, external forcings or internal processes, relative to the climate system. External forcings are reviewed here first.

The main external forcing of Holocene variability is the variation of the Earth's orbit, defined by three parameters: eccentricity, obliquity, and precession. These parameters are the main drivers of glaciation and deglaciations, as first theorized by Milankovitch in the 1920s (Berger, 1988). However, they have smaller impact on shorter time period such as the Holocene. For example, the eccentricity has not significantly changed in the last 10 000 years. By contrast, obliquity have decreased since its maximum 9500 years ago. Since seasons derive from the obliquity of the Earth, i.e. its rotation tilt compared to its orbit, the stronger obliquity during the early Holocene means that seasonality was more pronounced at that time than nowadays. As a result of the change in seasonality strength, the obliquity also impacts the amount of energy received at the equator in comparison to the poles.

Even more important is the change in precession. Nine thousand years ago, the Earth was at its closest to the Sun (perihelion) during northern hemisphere summer. Nowadays, the perihelion occurs during northern hemisphere winter. Hence, northern hemisphere seasonality was further enhanced during the early Holocene by changes in precession. During the mid-Holocene (5500 years ago), the perihelion occurred during northern hemisphere autumn. Interestingly, the precession also impacts the length of the season by a few days, since the Earth moves quickly around the Sun when at its perihelion (cf. third Kepler's law). Hence, some adjustments may be needed to compare the seasons at different periods in climate simulations (Joussaume and Braconnot, 1997).

The main impact of the precession and obliquity changes concerns the summer monsoon. The increased seasonality during the early Holocene strengthened the summer monsoon and the associated precipitation as demonstrated by proxy analysis (Clift and Plumb, 2008; Wang, 2006) and modelling studies (Kutzbach et al., 2007). By contrast, the progressive weakening of northern hemisphere seasonality throughout the Holocene led to a weakening

of the summer monsoon. Impacts on the winter precipitation are not clearly identified and are a matter of further study.

The change in orbital parameters is gradual and cannot directly explain climate variability on millennial and smaller timescales, without the effect of internal climate processes (cf. Section 1.4.3). Other external forcings can have an impact on shorter timescales, such as volcanic eruption.

Volcanic eruptions typically cool the Earth's surface in the following years while impacting the atmospheric circulation (Robock, 2000). A cluster of eruptions may have more lasting effects during one or several decades. Records of volcanic eruptions have been derived from ice cores in Greenland and the Antarctic (Castellano et al., 2005; Kobashi et al., 2017) and highlight that strong eruptions of Hekla (Iceland; Dugmore et al., 1995) and Cerro Blanco (Fernandez-Turiel et al., 2019, in the Andes;) occured close to the 4.2 event. Nonetheless, these proxies have large uncertainties, particularly for tropical eruptions which have a different climatic response than high-latitude eruptions (Oman et al., 2005). Eruptions may have dramatic human consequences. For example, it has been suggested that the explosion of Thera on Santorini around 3550 BP led to the collapse of the Minoan civilisation in Greece although as a result of a tsunami and earthquakes rather than due to a related change in the climate (Antonopoulos, 1992).

Finally, the Sun may exhibit variations of its intensity on various timescales. There exists some reconstructions across the Holocene, but with large uncertainties (Steinhilber et al., 2009; Vieira et al., 2011). The importance of solar variability for Holocene climate still remains controversial (Turner et al., 2016).

To summarise, orbital forcing explains well the multi-millennial Holocene trends observed in climate records, but the origin of multi-decadal variability and episodes like the 4.2 event is more complex. Particularly, they may not be externally forced and instead arise from internal process which are discussed further in the next section.

1.4.3 Internal processes and feedbacks

Variations of the climate on various timescales can also arise from internal processes, either independently or in response to an external forcing (i.e. a feedback). The most important factors are the atmospheric concentrations of carbon dioxide and methane, two key greenhouse gases in the glaciation-deglaciation oscillation. Outside of anthropogenic emissions, they are modulated by the carbon cycle through vegetation and the ocean at the timescale considered. Accurate records of greenhouse gas concentrations during the Holocene are available from ice cores (Flückiger et al., 2002), and show that carbon dioxide had a small rising trend between the early Holocene (260 ppm) and pre-industrial conditions (280 ppm). By contrast,

methane concentrations exhibited a low during the mid-Holocene (5000 BP; 550 ppb) while both early Holocene and the pre-industrial period were at similar level (700 ppb). As a result of different greenhouse gas concentrations, climate simulations suggest that the mean surface temperature of Earth was around 0.3 K lower during the mid-Holocene compared to pre-industrial conditions (PMIP4 models; Brown et al., 2020), although this is not supported by proxy data (Liu et al., 2014).

Vegetation and land cover can not only impact greenhouse gas storage and emissions, but also dust aerosol emissions from arid areas as well as evapotranspiration. Recent modelling studies have suggested that vegetation through both dust aerosol and evaporation may have been the main positive feedback explaining the Sahara greening in the early to mid-Holocene and its subsequent re-aridification (Pausata et al., 2016).

Changes in ocean circulation typically impact the multi-decadal to centennial variability of climate through, for example, variations of the Atlantic Meridional Overturning Circulation (AMOC). The AMOC was particularly coupled with the Laurentide ice sheet. Freshwater pulses from the melting ice sheet were responsible for transient collapses of the AMOC resulting in climatic events such as the Younger Dryas (Carlson, 2010) or the 8.2 event (Hoffman et al., 2012). However, the Laurentide ice sheet was already completely melted by about 6000 BP (Dyke et al., 2003) and freshwater pulses have not affected the AMOC since then.

The inter-annual timescale is influenced by coupled atmospheric-oceanic processes such as the El-Niño/Southern Oscillation (ENSO) which impacts both winter precipitation in the Indus River Basin (Dimri, 2013b; Yadav et al., 2010) and South Asian summer monsoon (Ashok et al., 2004). ENSO is thought to have varied in intensity and frequencies across the Holocene (Emile-Geay and Tingley, 2016; Eyring et al., 2016). The Indus River Basin is also affected by fluctuations in sea surface temperature in the Indian Ocean (e.g. the Indian Ocean Dipole, IOD; Ashok et al., 2004).

Finally, on even shorter timescales, weather is mostly affected by atmospheric processes. For example, winter precipitation is related to synoptic waves: the Western Disturbances (Dimri et al., 2015), which are themselves related on longer timescales to other atmospheric features, such as the jet-stream position or the North Atlantic Oscillation (NAO; Hunt et al., 2018a).

1.5 Thesis aim and approach

This thesis aims to identify the added-value of global climate models to climate records in order to characterise climate variability at the time of the Indus Civilisation (Section

1.1). This aim corresponds to an evaluation of climate model simulations. Precipitation is the key climate variable that affects human activities in the area (Section 1.3.2), either from direct rainfall on the plain, or snowfall in the upper mountains of the basin (Section 1.2.2). Hence, precipitation in the Indus River Basin as a whole is considered. Climate and weather phenomena can affect humans on a wide range of timescales. Figure 1.8 shows the corresponding timescales between human activities and various processes and forcings affecting precipitation in the Indus Basin. Processes internal to human society such as population growth, political upheaval, and agency are as important as processes internal to the climate system to explain their respective variability. Yet, human societies are also affected by precipitation variability on a wide range of timescale as highlighted in Figure 8. Hence, information is needed at various timescales, from the synoptic scale, the seasonal cycle, the inter-annual to multi-decadal variability and up to the multi-millennial Holocene trends, and each of these need to be evaluated in climate models.

In the following chapters, global climate models are first evaluated using present-day simulations in order to be able to compare them to instrumental observations, with a focus on the mean precipitation, the seasonal cycle and the processes explaining the synoptic to inter-annual variability. A good representation of these elements is key before investigating past precipitation and precipitation variability at larger timescale. Since this represents a too large scope for the study, this one is mostly limited to the area where both winter and summer precipitation overlap, in the upper part of the Indus River Basin, and with a focus on winter precipitation.

Chapter 2 will focus on determining the best observational datasets in terms of precipitation mean and variability at different timescales. This datasets will then be used as a

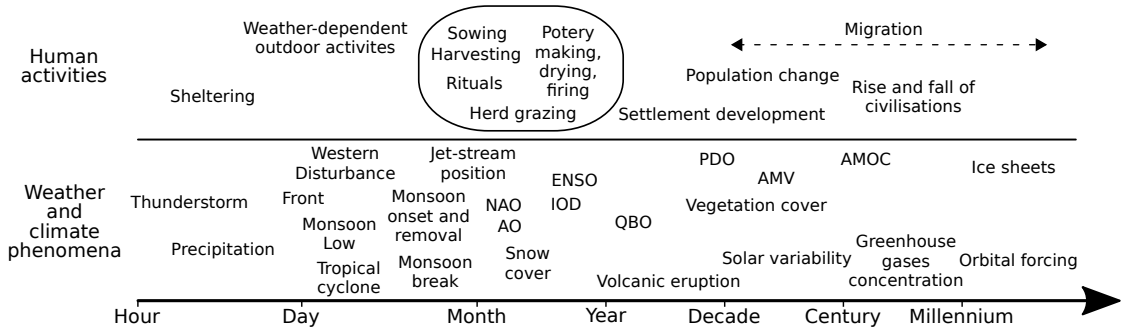


Fig. 1.8 Diagram of the correspondence between human activities and weather and climate phenomenons across timescales. The round shape correspond to seasonal human activities, with a similar timescale. (Acronyms not defined in text, QBO: Quasi-Biennal Oscillation, PDO: Pacific Decadal Oscillation, AMV: Atlantic Multi-decadal Variability)

reference for subsequent chapters. Chapter 3 will discuss the trigger of precipitation in the upper Indus Basin and its representation in several CMIP6 simulations. Chapter 4 will focus on characterising the link between winter precipitation and Western Disturbances which also helps to understand the precipitation seasonality. This knowledge will then be used in Chapter 5 to characterise the ability of the IPSL model with different setups across various timescales.

Chapter 2

Cross-validating precipitation datasets in the Indus River Basin

Preface

The content of this chapter, both in terms of methodology and results, has been published in Baudouin et al. (2020b). Co-authors C.A. Petrie and M. Herzog provided guidance and reviewed the paper. Only small modifications have been performed on the introduction and conclusion compared to the published version, so the chapter can fit in the thesis.

2.1 Introduction

Evaluating a climate model implies comparing its outputs to a reference, generally an observational dataset. This chapter will determine the most reliable precipitation dataset adapted to this purpose. A variety of observational datasets are available for the instrumental period. Yet, studies investigating the water cycle in the Indus River Basin have stressed the uncertainties inherent to observed precipitation (Dahri et al., 2018; Gardelle et al., 2012; Immerzeel et al., 2015; Singh et al., 2011; Wang et al., 2017). Therefore, a comprehensive comparison between observational datasets representing precipitation in the Indus Basin is needed.

Gridded products allow for a homogeneous spatial representation of precipitation at a river basin-scale for statistical purposes (Palazzi et al., 2013). They can be derived from rain gauges, satellite imagery or atmospheric models (e.g. reanalysis), but need validation to assess their quality. Most studies that validate precipitation products in Pakistan, India, or in the adjacent mountainous areas (Hindu-Kush / Karakoram / Himalayas) make use of rain

gauge data as a reference, either directly from weather stations (Ali et al., 2012; Ghulami et al., 2017; Hussain et al., 2017; Iqbal and Athar, 2018; Khan et al., 2014), or after gridding (Palazzi et al., 2013; Rajbhandari et al., 2015; Rana et al., 2015, 2017). However, some authors have pointed out that these reference datasets also suffer from limitations that could dramatically reduce correlation and increase biases, incorrectly lowering the confidence in the dataset validated (Ménégoz et al., 2013; Rana et al., 2015, 2017; Tozer et al., 2012).

The first issue of validating gridded precipitation products with rain gauge measurements is simply the uncertainty of the measurements. Beside the risk of corruption or missing values in the reporting process, it has been demonstrated that rain gauges can underestimate precipitation (Goodison et al., 1989; Sevruk, 1984). The main source of underestimation is wind-driven under-catchment that can reach up to 50% in case of snowfall (Adam and Lettenmaier, 2003; Dahri et al., 2018; Goodison et al., 1989; Wolff et al., 2015), but also includes wetting of the instrument, evaporation before measuring, and splashing out (WMO, 2008). Dahri et al. (2018) used the guidelines from the World Meteorological Organization (WMO) to re-evaluate the precipitation measured from hundreds of rain gauges in the upper part of the Indus Basin and found the underestimation to be between 1 and 65% for each station, and 21% basin wide. The second issue is the one of spatial representativeness. A rain gauge records a measurement at a specific location whereas in a gridded dataset, each value represents the mean over all the grid box. Thus, the two types of data have a different spatial representativeness. This discrepancy in representativeness increases when considering shorter timesteps and areas with strong heterogeneity such as mountainous terrains, which is especially impactful when studying extreme events. Some methods exist to quantify and tackle this issue (e.g. Habib et al., 2004; Tustison et al., 2001; Wang and Wolff, 2010).

Gridding methods are used to spatially homogenise point measurements and they also have limitations. Firstly, the specificity of the interpolation method can impact the result (Ensor and Robeson, 2008; Newlands et al., 2011). Secondly, the sparsity of the weather stations increases the uncertainties, which can range from 15 to 100% in areas with a low number of rain gauges (Rudolf and Rubel, 2005). This last point is especially problematic in the Indus Basin. For climatological purposes, the WMO has published guidelines for the density of rain gauges: from one station per 900 km^2 in flat coastal areas, to one every 250 km^2 in mountains (WMO, 2008). However, the Meteorological Department of Pakistan have recently published a 50-year climatology of precipitation for the country based on 56 stations, that is around one station per 15,000 km^2 (Faisal and Gaffar, 2012). Gridded rain gauge-based datasets rely on a similar density of observations in the Indus Basin (cf. Figure 2.1, Table 2.2). The situation in India is better as the Indian Meteorological Department produces a country-wide dataset of precipitation that is used for monsoon monitoring and

includes up to 6300 stations. This distribution makes around one station per 500 km, which is well within the WMO guideline. However, areas of lower density remain, especially in the western Himalayas and the Thar Desert, which are both in the Indus Basin (Kishore et al., 2016). Rain gauges are not only scarce in mountainous areas, but their location is also biased. In order to be accessible all year long, they are generally situated at the bottom of valleys, and these locations appear to be significantly drier than locations at altitude (Archer and Fowler, 2004; Dahri et al., 2018; Immerzeel et al., 2015; Ménégoz et al., 2013), which means that the interpolation method underestimates precipitation in the surrounding mountains.

There are a number of ways of overcoming the limitations of gridded rain gauge data, including the use of data derived from satellites and reanalyses. Satellite imagery can help to reduce both the lack and the heterogeneity of surface measurements. Satellite-based products generally make use of global infrared observations of cloud cover and microwave measurements along a swath (the narrow band where the observations are made as the satellite passes). However, their abilities over a heterogeneous terrain are more limited than over a flat and homogeneous one (Hussain et al., 2017; Iqbal and Athar, 2018; Khan et al., 2014). Moreover, these products still need rain gauges for calibration and are therefore dependent on the quality of station data.

Reanalyses of the atmosphere offer another way to estimate precipitation. Many valuable variables in a reanalysis are the result of the assimilation of observations with model outputs, but estimates of precipitation are, in most cases, a pure model product. That is, the precipitation is a forecast generated by the model used for the reanalysis, and is not constrained by direct observations in the way that other assimilated quantities are. Models are known to predict precipitation with difficulty and most studies highlight that precipitation from reanalyses is less reliable than that based on observations (Kishore et al., 2016; Rana et al., 2015). The reasons often invoked include discrepancies in spatial patterns and important model biases. However, recent progress in assimilation techniques has made it possible to integrate precipitation observations in the most recent reanalysis product (ERA5, Hersbach et al., 2018), and significant improvements are possible (e.g. Beck et al., 2019).

This study aims to better understand the quality and limitations of 20 precipitation datasets that are available for a study area encompassing the Indus Basin. Previous studies have investigated the strengths and limitations of precipitation datasets in this area (e.g. Ali et al., 2012; Hussain et al., 2017; Khan et al., 2014; Palazzi et al., 2013), but none has looked at such a large number of datasets nor at the most recent ones. Moreover, our method slightly differs, as we offer a cross-validation, thereby avoiding the problems that come from the selection of a unique reference. We cross-compare each of the datasets, identify their similarities and discrepancies, and using the diversity of data source and methods, assess

their strengths and weaknesses. After presenting the datasets selected for the study, we give a general description of the methods. The subsequent result section is split into four parts, which review, for the precipitation: i) the annual mean, ii) the seasonality, iii) the daily variability, and iv) the monthly and longer term variability. The final section concludes with the main results, the potential of the method, and future research priorities.

2.2 Data and Methods

2.2.1 Study areas

The Indus Basin extends across the north-westernmost part of the South Asian sub-continent, and is an area of various topographic features, as indicated in Figure 1.1. It is bounded from the north-east to the west by high mountain ranges, including the Himalayas, Karakoram, Hindu Kush and Sulaiman Ranges. To the south, the Indus River flows into the Arabian Sea. The eastern border is the most ambiguous as it extends into the flat dune-fields of the Thar desert. Much of the precipitation that falls in this extensive area evaporates before reaching the Indus River or the sea. It may also forms seasonal rivers, such as the Luni River, which has been included in the study area. This particular river flows into the Rann of Kutch, which is a flat salt marsh with complex connections with the Arabian Sea and the mouth of the Indus River (Syvitski et al., 2013), and is bounded to the west by the Aravalli Range. Although not strictly a part of the Indus watershed, it provides a clear and steady boundary for the study area. The total watershed considered for the study is represented by the outer blue contour line shown in Figure 2.1-A.

Precipitation amount varies across the basin as shown in Figure 2.1-A, as well as its seasonality. In the flat southern part, most of the precipitation occurs in July and August, under the influence of the South Asian summer monsoon propagating from the Indian Ocean and India, while during the rest of the year the basin remains dry (e.g. Ali et al., 2012; Khan et al., 2014; Rana et al., 2015). By contrast, the northern region is much more mountainous and encompasses a steep maximum of precipitation along the Himalayan front. This precipitation falls throughout the year, exhibiting a seasonal bi-modality explained by differences in circulation patterns (e.g. Archer and Fowler, 2004; Filippi et al., 2014; Palazzi et al., 2013; Singh et al., 2011). As in the southern part of the basin, a sharp peak in precipitation occurs in July-August related to the summer monsoon, but a second, broader peak also occurs in winter, from January to April, triggered by mid-latitude, extra-tropical Western Disturbances (Cannon et al., 2015; Dimri and Chevuturi, 2016; Hunt et al., 2018a).

Those differences in relief and precipitation seasonality and pattern suggest that the basin can be split into two distinct areas, along a line between 68.75°E-33.5°N and 77.5°E-30°N (inner blue contour in Figure 2.1-A), which broadly corresponds to the 100mm isohyet of winter precipitation (defined from December to March, cf. Figure 1.2-C). Thus, the northern part of the basin (hereafter the Upper Indus Basin or UIB, 595 000 km) includes the maxima of precipitation along the Himalayas and most of the winter precipitation, while the southern part (hereafter the Lower Indus Basin, LIB, 785 000 km) is characterised by a single wet season during summer, as wintertime precipitation is negligible (Figure 2.2).

2.2.2 Datasets

2.2.2.1 Rain gauge data

We have selected five commonly used and one newly available gridded dataset based only on rain gauge data. These are the first six datasets presented in Table 2.1. The mean number of stations used in the two study areas are available for five of the datasets and presented in Table 2.2. The Asian Precipitation Highly Resolved Observed Data Integration Towards Evaluation of water resources (APHRODITE; Yatagai et al., 2012) was developed by the Research Institute for Humanity and Nature (RIHN) and the Meteorological Research Institute of Japan Meteorological Agency (MRI/JMA). Specific to Asia, it is one of the best datasets available for the area (Rana et al., 2015), both in term of resolution (0.25° and daily, it includes a large number of rain gauges; Table 2.2) and because it covers an extended period (over 50 years). However, it does not provide data after 2007. A new dataset has been issued in 2019 from the same institute extending the period covered up to 2015 and using a new algorithm (APHRODITE-2), though its quality has not yet been investigated. Covering the whole twentieth century at a monthly resolution, the Global Precipitation Climatology Center monthly dataset (GPCC-monthly; Schneider et al., 2018) is widely used in climatology and for calibration purposes (e.g. satellite-based datasets, Table 2.1). GPCC-daily (Ziese et al., 2018) offers a better temporal resolution (daily), but at a lower spatial resolution and has a much-reduced time coverage compared to GPCC-monthly. It uses a smaller number of rain gauges (Table 2.2), but is constrained by GPCC-monthly. The precipitation dataset from the Climate Research Unit (CRU; Harris and Jones, 2017) has a similar resolution and time coverage as GPCC-monthly. We also selected another daily dataset from NOAA's Climate Prediction Center (CPC; Xie et al., 2010). Although CPC uses a lower number of rain gauges compared to APHRODITE (Table 2.2), its availability extends to the present with near real time updates, which means that it can be used for calibrating other near real time products (e.g. CMAP in Table 2.1 and MERRA2 in Table 2.3).

2.2.2.2 Satellite data

Various satellite-based gridded precipitation products are available, but we have only selected datasets providing data from 1998, to ensure a long enough common period with the rain gauge-based datasets (the common period reaches 10 years due to APHRODITE ending in 2007). Four were eventually selected (last four datasets in Table 2.1). The Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al., 2007) is the most widely used satellite-based datasets. It has the highest temporal and spatial resolution of the selection (sub-daily, and 0.25° like APHRODITE and GPCC-monthly) and includes a large diversity of satellite observations. We also selected the daily product from the Global Precipitation Climatology Project (GPCP-1DD; Huffman and Bolvin, 2013) as well as the monthly product issued by the same group (GPGP-SG Adler et al., 2016). All three of these datasets (TMPA, GPCP-1DD, and GPGP-SG) use GPCC for calibration, which could introduce some similarities. By contrast, the last dataset included, CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997), uses CPC for calibration. It has the same time coverage and resolution as GPCP-SG. This version does not include reanalysis data, to simplify the analysis.

2.2.2.3 Reanalysis data

Unlike the observation datasets, reanalysis data can be quite different from one another. They generally use their own atmospheric model and assimilation scheme, and the type and number of observations assimilated can vary. Table 2.3 shows the ensemble of the ten reanalysis datasets that have been used in this study. The four reanalyses of the latest generation are, from most recent to oldest: ERA5 (Hersbach et al., 2018) from the European Centre for Medium-Range Weather Forecasts (ECMWF), the Modern Era Retrospective-analysis for Research and Applications version 2 (MERRA2; Gelaro et al., 2017) from the NASA, the Japanese 55-year Reanalysis (JRA; Kobayashi et al., 2015) from the JMA, and the Climate Forecast System Reanalysis (CFSR; Saha et al., 2010, 2014) from the National Center for Environmental Prediction (NCEP). These are still regularly updated, they all include the latest observations from satellites and cover the full satellite era from at least 1980. JRA goes back to 1958, when the global radiosonde observing system was established. ERA5 currently starts in 1979 but future releases are expected to extend this back to 1950.

In terms of technical differences, ERA5 uses a more complex assimilation scheme than the others reanalysis (4DVAR), which allows for better integration of the observations. It is also the only one that assimilates precipitation measurements. MERRA2 also uses observations, but takes them from a gridded dataset (CPC) and only uses them to correct

Table 2.1 Observational datasets of precipitation selected for this study, derived from rain gauges or satellites

Name	Version	Time coverage	Time resolution	Spatial resolution	Based on	Reference
APHRODITE	V1101	1951-2007	Daily	0.25°	Rain gauge only	Yatagai et al. (2012)
APHRODITE-2	V1901	1998-2015	Daily	0.25°	Rain gauge only	
CPC	V1.0	1979 (monthly) / 1998 (daily) -2018	Daily	0.5°	Rain gauge only	Xie et al. (2010)
GPCC-daily	V2	1982-2016	Daily	1°	Rain gauge and GPCC-monthly	Ziese et al. (2018)
GPCC-monthly	V8	1891-2016	Monthly	0.25°	Rain gauge only	Schneider et al. (2018)
CRU	TS4.02	1901-2017	Monthly	0.5°	Rain gauge only	Harris and Jones (2017)
TMPA	3B42 V7	1998-2016	3-hourly	0.25°	GPCC, satellites	Huffman et al. (2007)
GPCP-1DD	V1.2	1996-2015	Daily	1°	GPCC, satellites	Huffman and Bolvin (2013)
GPCP-SG	V2.3	1979-2018	Monthly	2.5°	GPCC, satellites	Adler et al. (2016)
CMAP	V1810	1979-2018	Monthly	2.5°	CPC, satellites	Xie and Arkin (1997)

Table 2.2 Number of stations used on average for the rain-gauge-based datasets (except CRU for which this information was not directly available), per time step, for the two study areas, and over the period 1998-2007.

Datasets	UIB	LIB
APHRODITE	55	48
APHRODITE-2	88	65
CPC	15	21
GPCC-daily	11	16
GPCC-monthly	35	33

the precipitation field before analysing the atmospheric impact on the land surface; this changes land surface feedbacks on the atmosphere. CFSR is an Ocean-Atmosphere coupled reanalysis, that is, the sea surface is modelled and provides feedback to the atmospheric model, instead of being prescribed by an analysis from observations. ERA5 and MERRA2 are the most recent of the reanalysis datasets to be published, and not many studies have looked at the improvement from their predecessor, respectively ERA-Interim (Dee et al., 2011) and MERRA1 (Rienecker et al., 2011). Both have stopped being updated or will very shortly, but they are included in this study for comparison purposes.

Reanalyses for the whole twentieth century have also been produced, but to retain the homogeneity of the type of observations assimilated they only include surface observations. The twentieth century reanalysis from NCEP (20CR; Compo et al., 2011), only assimilates surface pressure, but more recently, the ECMWF produced ERA-20C (Poli et al., 2016), which has surface wind assimilated along with surface pressure.

We have also made use of older generation reanalysis datasets that are still being updated, including: the NCEP/NCAR reanalysis (NCEP1; Kalnay et al., 1996) and the NCEP/NDOE reanalysis (NCEP2; Kanamitsu et al., 2002). Both are useful to quantify the progress in reanalysis systems as well as to compare them with more observation-limited century long reanalyses.

2.2.3 Methods

For each dataset, the time series of precipitation are averaged over the two study areas (UIB and LIB) and calculated at a monthly resolution, and daily if possible. The datasets have different spatial resolution which causes a problem when calculating the precipitation averages over the study areas. Simply selecting the cells whose centre is within these areas leads to small biases in the extent of the region considered. These biases are reduced by

Table 2.3 Datasets of precipitation selected for this study, derived from reanalysis

Name	Time coverage	Spatial resolution	Remarks	Reference
ERA5	1979-2018	0.25°	4DVAR, precipitation assimilated	Hersbach et al. (2018)
ERA-Interim	1979-2018	0.75°	4DVAR	Dee et al. (2011)
JRA	1958-2018	0.5°		Kobayashi et al. (2015)
MERRA2	1980-2018	0.5° / 0.625°	Correction of the precipitation with CPC for land interaction. Assimilate aerosol observations	Gelaro et al. (2017)
MERRA1	1979-2010	0.5° / 0.66°		Rienecker et al. (2011)
CFSR	1979-2018	0.5°	Coupled reanalysis (atmosphere, ocean, land, cryosphere). Same analyses as MERRA1. Version 2 starting in 04/2011	Saha et al. (2010, 2014)
NCEP2	1979-2018	1.875°	Fixed errors and updated model since NCEP1 No satellite radiance assimilated	Kanamitsu et al. (2002)
NCEP1	1948-2018	1.875°	No satellite radiance assimilated	Kalnay et al. (1996)
20CR	1871-2012	1.875°	Assimilate surface pressure only	Compo et al. (2011)
ERA-20C	1900-2010	1°	Assimilate surface pressure and marine wind only	Poli et al. (2016)

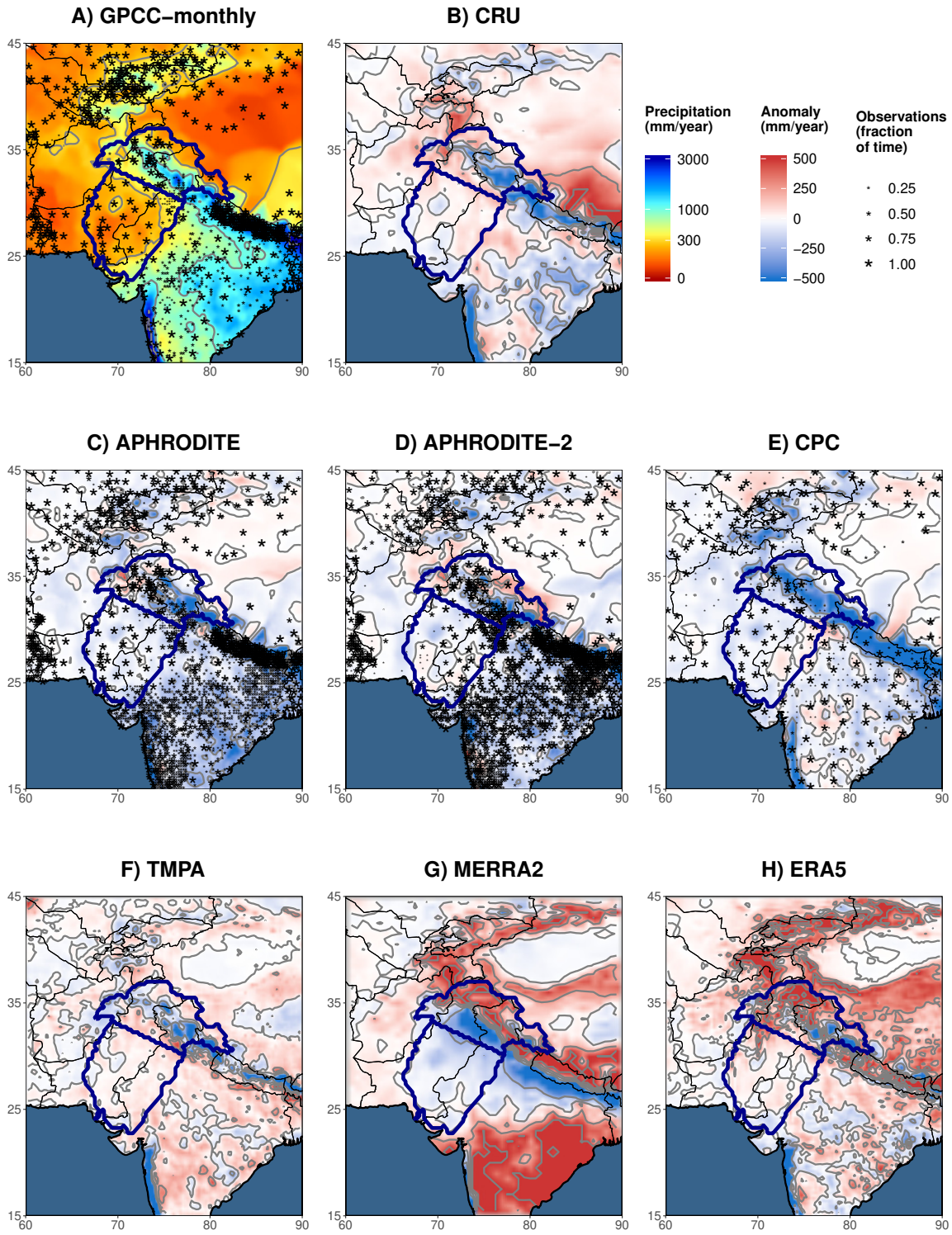

A) GPCC–monthly
B) CRU
C) APHRODITE
D) APHRODITE–2
E) CPC
F) TMPA
G) MERRA2
H) ERA5
Precipitation (mm/year)
3000
1000
300
0
Anomaly (mm/year)
500
250
0
−250
−500
Observations (fraction of time)
0.25
0.50
0.75
1.00
45
35
25
15
60
70
80
90

Fig. 2.1 (Previous page) Map of annual mean precipitation for different datasets. The annual mean is computed over the period 1998-2007. GPCC monthly (A) is used as a reference to compute the anomaly for the other datasets (B to H). The grey lines are the isohyets whose level corresponds to the labels in the legend. The boundaries of the two study areas are displayed in dark blue on each map. The stars mark the grid cells that include at least one gauge observation. The size of the stars represents the number of time steps with at least one observation over that cell, relative to the total number of time steps needed to compute the annual mean (120 for A, 3652 for C,D and E). This information was not available for CRU (B) nor ERA5 (H), and does not apply to the satellite-based TMPA (F), and MERRA2 (G).

---

bi-linearly interpolating all data to a 0.25° grid, common to APHRODITE, APHRODITE-2, and GPCC-monthly. This choice is further discussed in Section 2.3.1.1.

The analysis is performed over the 10-year period from 1998-2007, which is common to all datasets, except when analysing the trends and inter-annual to decadal variability, for which we use all data available. We focus on the two wet seasons of the UIB. Summer is defined from June to September, which matches the monsoon precipitation peak. Winter is defined from December to March. This fits the snowfall peak rather than the precipitation peak, but makes it possible to focus on the issue of snowfall estimation (Palazzi et al., 2013). In the LIB, we use the same definition for summer, but winter is not analysed, as it is a dry season.

We first compare the mean and seasonal cycle of each dataset in Sections 2.3.1 and 2.3.2. For quantitative statements, we use GPCC-monthly as a reference. However, in Section 2.3.1.3, we use the precipitation dataset from Dahri et al. (2018) as reference instead. This dataset cannot be used in other parts of the study, as it is limited to one part of the UIB, and only provides annual means.

Then, in Section 2.3.3 we compare the daily variability of the precipitation using the Pearson correlation. The correlation significance is discussed at the 95% probability level. To reduce the impact of unusually large rainfall events when investigating trends in daily variability (cf. Section 2.3.3.4), we use the Spearman correlation. Lastly, in Section 2.3.4, other time scales of variability of the precipitation are investigated: monthly, seasonal, inter-annual, and decadal, still using the Pearson correlation at the 95% confidence interval.

2.3 Results

2.3.1 Annual mean

2.3.1.1 Differences between rain gauge-based datasets

Annual mean precipitation in both study areas and for each dataset are given in Table 2.4 (last two columns). We first focus on the rain gauge-based datasets (upper part of the table). Spatial pattern differences are shown in Figure 2.1-A to E.

First, we should mention that the bi-linear method we use to interpolate each dataset to the same grid (cf. Section 2.2.3) leads to some differences between datasets. The two GPCC products can be used to evaluate the impact of our interpolation method, as they have a different spatial resolution but uses the same monthly climatology. Hence, the small underestimation of GPCC-daily compared to GPCC-monthly (about 1% in the UIB and 5% in the LIB) is related to the interpolation method. However, these differences are small enough to justify the use of our method.

More generally, annual mean differences can be explained by methods and data that each dataset uses. Particularly, the interpolation of station measurements to a grid differs from one dataset to the other. APHRODITE's interpolation method, for instance, considers the orientation of the slope to quantify the influence of nearby stations. This greatly reduces the amount of precipitation falling in the inner mountains compared to GPCC-monthly. An example of this pattern is evident to the north of the Himalayas where only very few observations exist (Figure 2.1-D; Yatagai et al., 2012). In CRU, the interpolation method (triangulated linear interpolation of anomalies; Harris et al., 2014) seems to smooth areas of strong gradients such as near the foothills of the Himalayas (Figure 2.1-B). This smoothing might explain a slightly drier UIB, and slightly wetter LIB, compared to GPCC-monthly (Table 2.4).

Differences can also be explained by the dramatic change in location and number of stations used to compute the statistics (Figure 2.1-A, C, D, and E, Table 2.2). For example, CPC is by far the driest dataset in the UIB and the second driest in the LIB. This is likely related to the low number of observations it includes, leaving vast areas with no or very few observations, including the wettest regions (Figure 2.1-E). However, there is no linear relationship between precipitation amount and number of observations. GPCC-daily includes the lowest number of observations, but this does not impact its climatology, because the climatology is derived from GPCC-monthly. On the contrary, APHRODITE comprises a much higher number of observations than others datasets, but remains much drier than GPCC-monthly (about 20% drier in both study areas).

Table 2.4 Mean annual and seasonal precipitation (in mm) falling over the two study areas, for the period 1998-2007. Winter is defined from December to March and summer from June to September. The first ten datasets are observations, the second ten are reanalyses.

Datasets	UIB			LIB
	Winter	Summer	Annual	Annual
APHRODITE	154	237	484	198
APHRODITE-2	179	272	555	223
CPC	98	200	355	216
GPCC-daily	201	297	607	243
GPCC-monthly	201	301	613	255
CRU	166	281	565	267
TMPA	156	298	555	286
GPCP-1DD	161	305	569	317
GPCP-SG	167	309	583	325
CMAP	273	307	696	279
ERA5	280	380	828	300
ERA-Interim	289	445	931	305
JRA	299	325	810	586
MERRA2	265	310	724	177
MERRA1	205	267	598	355
CFSR	282	214	656	162
NCEP2	274	259	703	276
NCEP1	372	343	915	239
20CR	244	319	746	116
ERA-20C	175	276	551	175

Yatagai et al. (2012) pointed out that differences in quality checks compared to the other datasets might explain APHRODITE's dry bias. They noted that APHRODITE partly relies on GTS data that are sent in near real time to the global network, with the risk of misreporting. This risk particularly concerns misreported zero values, which are hard to detect and lead to a dry bias. The large dry bias seen in CPC data might be associated to the same issue, since CPC is entirely based on GTS data. In GPCC-monthly (and daily), only stations with at least 70% of data per month are retained (Schneider et al., 2014), while in CRU this number is 75% (Harris et al., 2014). Thus, limiting the analysis to the most reliable weather stations can help build confidence in recorded total precipitation amount.

Interestingly, APHRODITE-2 is more than 10% wetter than APHRODITE in both study areas. Several changes have been performed in the methodology: quality control of extreme high values, station-value conservation after interpolation, merging daily observation with

different definitions of End of Day time (cf. Section 2.3.3.1), and an updated climatology. However, the difference in mean precipitation is most likely related to the change in observations from rain gauges. Although APHRODITE-2 comprises more observations basin-wide, this increase mainly occurs over the Indian territory, whereas Pakistan is presented with fewer precipitation measurements, especially in the dry southern central part (Figure 2.1-D). This decrease in observations in a drier area can reasonably explain the increase in mean precipitation in the LIB. In the UIB, the increase is mainly due to the inclusion of measurements from one isolated weather station in the northernmost part of the area.

2.3.1.2 Considering satellite and reanalysis datasets

We now consider satellite-based datasets (middle part of the Table 2.4). In the UIB, CMAP stands out as being the wettest observational datasets, 13% wetter than GPCC-monthly. By contrast, the other three (TMPA, GPCC-1DD, GPCP-SG), are drier than GPCC-monthly (between 10 and 5%), despite being calibrated by this GPCC-monthly. In the LIB, all satellite-derived datasets are wetter than the rain gauge products (between 10 and 30% more than GPCC-monthly). The complexity of the algorithm used to produce the satellite-based datasets makes determining the reasons for their differences with each other or with rain gauge products difficult. According to previous studies, their ability to represent precipitation over rough terrain is limited (e.g. Hussain et al., 2017). In fact, Figure 2.1-F shows that the strongest differences between TMPA and GPCC-monthly occurs near mountain ranges, such as the UIB. In contrast, precipitation estimates over flat terrain with sparse observations and mostly convective precipitation benefit from satellite observations (Ebert et al., 2007). This suggests that the higher precipitation mean of the satellite-derived datasets for the LIB is possibly due to better consideration of locally higher precipitation rates during convective events.

The annual mean precipitation in reanalysis datasets is listed in the lower part of Table 2.4. In the LIB, the range of values is very high: the wettest dataset, JRA, is five times wetter than the driest dataset, 20CR. This range shows the significant difficulties for reanalyses to represent precipitation in an area were convection dominates. Among the most recent reanalyses, ERA5 has the closest estimates of precipitation to the observational datasets, yet above the estimates from rain gauges. Figure 2.1-H suggests that these wetter conditions mainly comes from the north-western edge of the Suleiman range, an area with sparse precipitation observation (cf. Figure 2.1-A), therefore increasing confidence in ERA5 estimation. The two twentieth century reanalysis (20CR and ERA-20C) are amongst the driest reanalysis datasets, suggesting that their models have difficulties to propagate the monsoon precipitation into the LIB region, when only surface observations are assimilated.

Lastly, MERRA2 exhibits a severe drop of precipitation compared to the previous version, MERRA1. Summer monsoon precipitation is known to be strongly affected by surface moisture content, especially in flat areas like the LIB (Douville et al., 2001; Pathak et al., 2017). MERRA2 uses CPC data to constrain the precipitation flux at the surface. Due to the dry bias of CPC, soil moisture is reduced for most of India (Figure 3 in Reichle et al., 2017), explaining the drop in precipitation.

For the UIB, the most striking features is that all reanalysis datasets except MERRA1 and ERA-20C predict higher precipitation amounts than GPCC-monthly, about 20% higher on average. In the following we investigate whether that this difference can be explained by an underestimation of rain gauge measurements.

2.3.1.3 Impact of rain gauge biases in mountainous terrains

Rain gauge measurements are known to potentially underestimate precipitation and particularly snowfall (Goodison et al., 1989; Sevruk, 1984). This is an important issue for mountainous region such as the UIB. However, among the six rain gauge-based datasets, only GPCC's products consider a correction of the data. Based on a study by Legates and Willmott (1990), a correction factor, which depends on the month, is applied at each grid cell. Most of these factors vary between 5 and 10% in the UIB (Figure 4 in Schneider et al., 2014), and explain why GPCC's products are wetter than most of the other rain gauge-based datasets. Recently, Dahri et al. (2018, hereafter Dahri2018) compiled the measurements from over 270 rain gauges in the UIB and adjusted their values to undercatchment, following WMO guidelines. They found a basin-wide adjustment of 21%, but this varies from 65% for high altitude stations, to around 1% for the stations in the plains.

The Dahri2018 dataset has both the advantage of considering a very large number of observations and correcting rain gauge measurements. However, its result is based on a study area somewhat smaller than the UIB region presented here, and only covers the period from 1999 to 2011. For comparison purpose, we recomputed the annual mean of several of the most recent and highest resolution datasets to fit these definitions (Table 2.5).

Table 2.5 shows that none of the observational datasets is able to reproduce the Dahri2018 precipitation estimates. They all have a dry bias, from 30% for TMPA, to 10% for GPCC-monthly. Furthermore, APHRODITE-2 and TMPA even underestimate the unadjusted value of Dahri2018, which suggests that the underestimation is not only related to rain gauge measurements, but also to the representation of the spatial pattern. By contrast, GPCC-monthly is 7% higher than the Dahri2018 unadjusted value, which corresponds to the correction factor used in GPCC. This suggests that the unadjusted values in both datasets are very close, and highlights the quality of GPCC. Nevertheless, we also found discrepancies in

Table 2.5 Mean annual precipitation (in mm) for various datasets over the study area defined in Dahri et al. (2018) for the period 1999-2011. Both adjusted and unadjusted values (the latter in parenthesis) from Dahri et al. (2018) are reported in the second line

Datasets	Revised UIB
Dahri2018	697 (574)
APHRODITE-2	548
GPCC-monthly	612
TMPA	480
ERA5	835
JRA	827
MERRA2	929
CFSR	783

the spatial patterns between GPCC-monthly and Dahri2018. Particularly, in the northernmost part of the UIB region, in the Karakoram range, GPCC-monthly exhibits lower precipitation means than Dahri2018, which cannot be explained by the difference in correction factors between the two datasets alone. The nearest stations used in GPCC-monthly are all located in the dry and more accessible Indus River valley to the south of the mountain range (Figure 2.1-A). Those drier conditions extend to the north due to the interpolation method used by GPCC, while Dahri2018 includes station measurements with wetter conditions than in the valley. This difference illustrates the impact of biased weather station locations mentioned in the introduction and in several other studies (e.g. Archer and Fowler, 2004; Immerzeel et al., 2015; Ménégoz et al., 2013).

Still in the Karakoram range, Figures 2.1-G and H show that MERRA2 and ERA5 are wetter than GPCC-monthly, and therefore closer to Dahri2018. However, spatial discrepancies remain. Particularly, the maximum of precipitation in MERRA2 is shifted to the North, which leads to important biases when averaging on the Dahri2018's study area. Our study area, which does not overlap with the highest precipitation rates, is less affected by shifts and is better fitted to compare the large scale precipitation patterns. Nevertheless, the four selected reanalysis datasets in Table 2.5 overestimate the Dahri2018 adjusted values, by 20% on average. This suggests that part but not all of the differences between reanalyses and observational data can be explained by biases from the latter. This overestimation of modelled precipitation in reanalyses for the UIB is corroborated by previous studies (e.g. Palazzi et al., 2015).

To conclude, all rain gauge-based datasets suffer from an underestimation of annual mean precipitation for the UIB when compared to Darhi2018. This results from biases in rain gauge locations and measurements. Quality control and interpolation methods also impact

precipitation amount in both parts of the basin. Satellite observations probably improve precipitation estimates in flat areas with sparse observations. However, they cannot correct observational biases since they use them for calibration, and biases remain unchanged or even amplify for the UIB. Reanalyses do not include rain gauge measurement, except for ERA5 and MERRA2, and are therefore not affected by observational biases. However, model biases can also be significant as suggested by the spread of the annual mean precipitation values. Reanalyses tend to be wetter than observational datasets in the UIB, which is partly explained by the underestimation of the observations. Lastly, all datasets suffer from spatial discrepancies, which are detrimental to small-scale comparisons, especially near mountains, but justify our choice to use a larger study area.

2.3.2 Seasonal cycle

The seasonal cycle of precipitation for each dataset is presented in Figure 2.2. Analysing the seasonality is particularly interesting in the UIB, as it is characterised by two wet seasons. The mean precipitation of each season is presented in Table 2.4 (second and third column). The rain gauge-based datasets exhibit a very similar seasonality for both study areas. In the UIB, the maxima of precipitation occur in February and July, the minima in May and November. The differences between the datasets vary little from one month to another, which suggests that the causes of the differences identified in the previous section (e.g. misreporting, station location and number, interpolation method) are independent of the seasonality. The satellite-based datasets represent the summer precipitation almost exactly as GPCC-monthly. The annual mean differences are explained by biases during the winter season, which suggests that winter precipitation is more difficult to estimate for those datasets.

The reanalyses represent the dry and wet seasons of the UIB, but with a larger spread than in the observations and some differences in seasonal cycle (Figure 2.2-B). On average, winter precipitation is 30% higher than in GPCC-monthly, with the notable exception of ERA-20C (Table 2.4). Those wetter conditions also extend to the surrounding drier months: April/May and October/November. However, the mean summer precipitation in reanalyses is not significantly different from GPCC-monthly (Table 2.4). Only ERA-Interim stands out with a wet summer precipitation bias, mainly in the north-west corner of the UIB, a bias partly corrected in ERA5 (Figure 2.1-H). The winter wet bias is not surprising after the comparison with the Dahri2018 dataset in the Section 2.3.1.3. Indeed, Dahri2018 found that the most important rain gauge underestimations happen in winter when precipitation mostly falls as snow. More interestingly, we found that the latest reanalyses (ERA5, JRA, MERRA2, and CFSR) represent winter precipitation in similar ways. We haven't been able to investigate

the seasonality of the Dahri2018 dataset, but we suggest that the latest reanalyses better represent winter precipitation than the observational datasets.

We noted another discrepancy in seasonality between a majority of the reanalyses and the observations for the UIB: a delay of the summer precipitation starting from the pre-monsoon season (Figure 2.2-B). The observations show that May is the driest month of that season, followed by a sharp increase in precipitation in June. Only ERA5, ERA-Interim, and MERRA1 reproduce this behaviour. In contrast, NCEP2 and CFSR are much drier in June than in May. For other reanalyses, precipitation during May and June are comparable. This delay continues into the summer monsoon period: while the observations clearly show a wetter July than August, this is only the case for ERA5, ERA-Interim, and both MERRA reanalyses. A similar delay can be found over the Ganges plain and along the Himalayas, which suggests wider uncertainties on the monsoon propagation in the reanalyses. By contrast, no such delay is found in the LIB, despite the large uncertainty on the amount of precipitation (Figure 2.2-D).

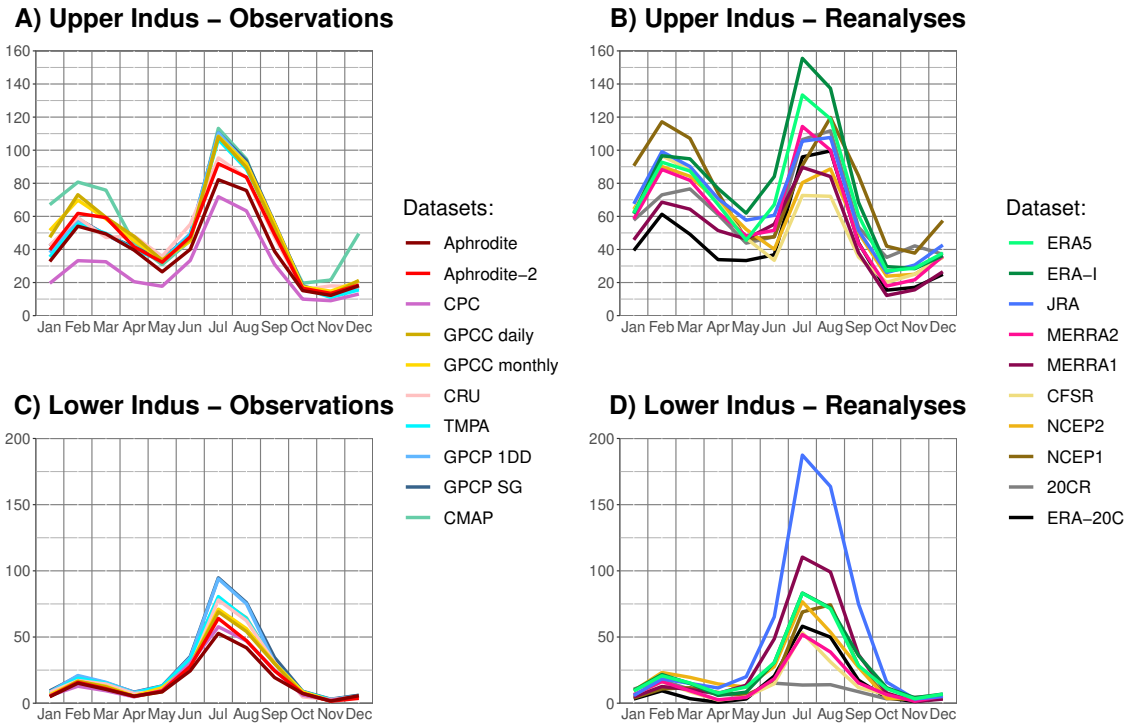


Fig. 2.2 Monthly mean of precipitation, over the period 1998-2007, representing the seasonal cycle. Results are split between UIB (A and B) and LIB (C and D) as well as observation datasets (A and C) and reanalysis (B and D).

2.3.3 Daily variability

2.3.3.1 Lag analysis

Investigating the daily precipitation variability helps to better quantify the quality of each dataset. Before computing the daily correlation, we checked for possible lags between the datasets. Lags can have different origins. The first is the accumulation period considered for the rain gauge measurements. CPC documentation (Xie et al. 2010) points out that the official period is different from one country to another (in our case, Afghanistan, Pakistan, and India all use different periods, or "End of Day time": 00hUTC, 06hUTC, 03hUTC, respectively), which could impact precipitation estimates. Neither GPCC-daily nor APHRODITE documentation mention this issue, while a specific effort has been made to homogenise all observations in APHRODITE-2. Secondly, the TMPA algorithm uses the 00h imagery for the following day accumulation, and therefore, could be more representative of an accumulation starting at 22:30h UTC (Huffman et al., 2007). Thirdly, biases in the daily cycle are possible in the reanalyses.

Our main finding relates to CPC. Figure 2.3 shows the daily correlation year per year of CPC against APHRODITE and MERRA2, for two lags: 0h and -24h (previous day for CPC). We found that the two lags switch their behaviour somewhere around 1997/1998, which we interpret as an error in the data processing for CPC. That is, in CPC before 1998, precipitation values correspond to those for the following day. This should not have an important impact on monthly and longer accumulations, but we limited the daily analysis of CPC to the period from 1998 to 2018. Moreover, similar errors might have happened earlier during the 1980s as the curves in Figure 2.3 come closer or invert again. This error also propagates to the corrected precipitation of MERRA2. That is, before 1998, the land surface in the model receives the precipitation of the following day. Theoretically, this could enhance precipitation by increasing surface moisture supply before the precipitation actually falls. However, we have not been able to find a significant change before and after 1998. The error has been reported to NOAA's CPC.

Possible differences in the End of Day times of the observational datasets are investigated using the sub-daily resolution of TMPA. We compute TMPA daily accumulation with different End of Day time and determine which one maximises the correlation with the other observational datasets. APHRODITE and CPC (after 1998) maximise the correlation with TMPA when for the latter an End of Day at 03h UTC is used. This behaviour suggests that both CPC and APHRODITE are more representative of an accumulation period ending at 03hUTC, influenced by the Indian rain gauge network. APHRODITE-2 successfully

corrected this delay, maximising correlation with TMPA for a End of Day at 00hUTC, like GPCC-daily.

A similar analysis can be performed for the reanalyses, to investigate the possibility of a delay in the daily cycle of the precipitation. We found that most reanalyses have a negligible ($\leqslant$3h) delay with TMPA. However, the reanalyses of the twentieth century have a different behaviour: both have a +12h delay. For those two, only surface observations are assimilated. It is possible that 12h is the time needed by the troposphere to adjust to those surface constraints. Worst constraint on convection could also impact the timing of precipitation. However, numerical models are known to trigger convection too early which is contrary to the +12h delay found here.

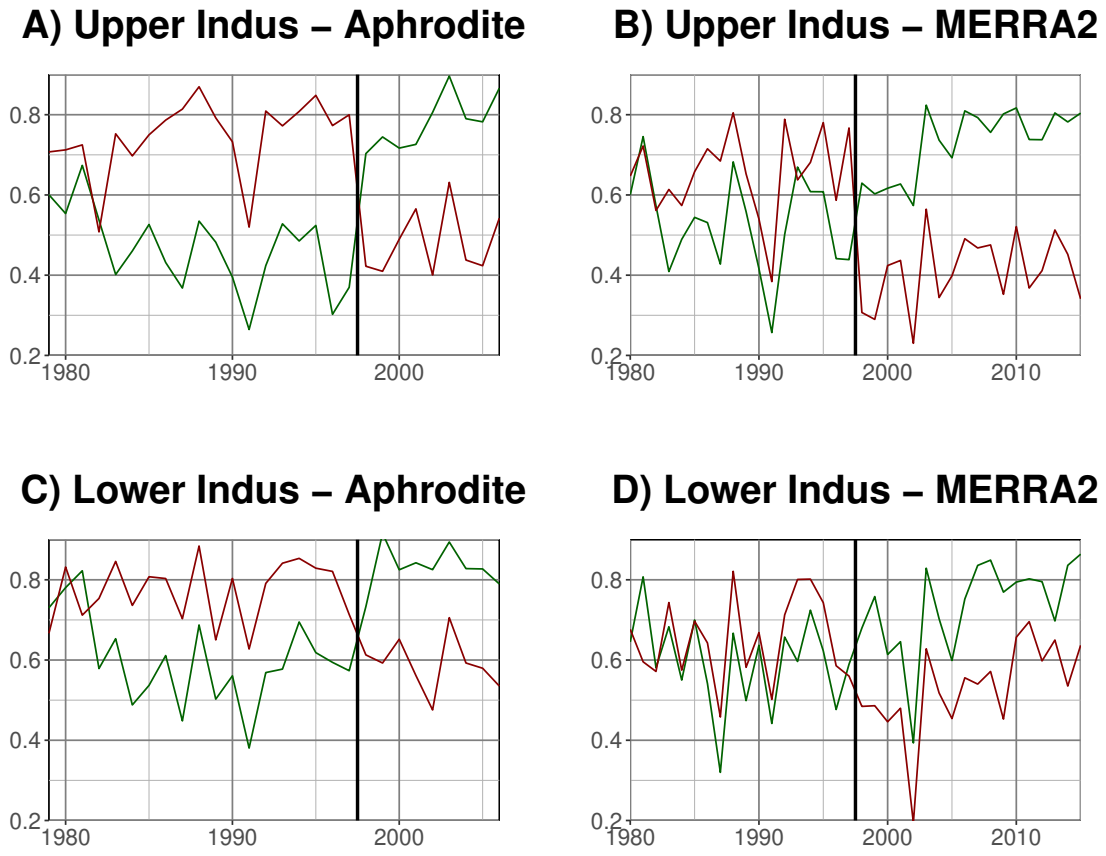


Fig. 2.3 Daily correlation, per year, between CPC and APHRODITE (A and C), and MERRA2 (B and D) for both UIB (A and B) and LIB (C and D). The green line is the correlation between the same days in each dataset. For the red line, the previous day of CPC is used instead. The black vertical line is the start of the year 1998, around where the main error should be.

Finally, we decided to take the accumulation period starting at 00h for all sub-daily datasets. Indeed, it is not straightforward to correct the delay in APHRODITE or CPC for instance, since only a daily resolution is available. Moreover, the correlation coefficients are not too importantly affected by those sub-daily lags.

2.3.3.2 Cross-validation

We now start the comparison of the daily variability between each dataset. Particularly, we aim to understand whether the co-variability exhibited between datasets is coming from the use of a common method or data source, or from a good representation of the precipitation variability. All datasets are estimates of precipitation, but they use different methods and input data to achieve this (cf. Section 2.2.2). If two datasets share a similar method or data source, this can at least partly explain the co-variability between the datasets. If, on the contrary, the two datasets are independent, then the co-variability they share is most likely due to the precipitation signal they estimate. As a consequence, the higher the correlation between two independent datasets, the better is the estimate of precipitation of both datasets.

Table 2.6 presents the daily correlation of precipitation between the different datasets, for the UIB. The upper part of the table focuses on the cross-correlation between the observational datasets. The highest correlation coefficient, almost 0.9, is between TMPA and GPCP-1DD, showing how close those two datasets are, likely due to the satellite observations they have in common and the similarity of retrieval procedures (Palazzi et al., 2013; Rahman et al., 2009; Rana et al., 2017). The rain gauge-based datasets APHRODITE, CPC, and GPCC-daily have also a high correlation between one another of about 0.8. The two versions of APHRODITE are even closer, due to their similarities of conception. When comparing GPCP-1DD and TMPA's correlation coefficients using the rain gauge-based datasets as reference, it turns out that the TMPA coefficients are systematically significantly higher than those for GPCP-1DD (at the level 95%). That is, TMPA variability is closer to the rain gauge-based datasets than GPCP-1DD is. It could be either because TMPA includes more information from the rain gauge measurements than GPCP-1DD or because it has better quality (better algorithm, better data source). Similarly, we note that APHRODITE and APHRODITE-2 have significantly higher correlation with the satellite-based datasets than CPC and GPCC-daily do.

In the lower part of Table 2.6, the correlation between the reanalyses and the observational datasets are about as high as between the observational datasets, suggesting that reanalyses are as good as observational datasets in representing the daily variability. Moreover, precipitation from reanalysis and observational data are independent from each other, in the sense that they do not share the same data source (except ERA5, which assimilates precipitation observations, and MERRA2, which integrate CPC data; the two need to be treated separately). Hence, the

Table 2.6 Daily correlation between different datasets, in the UIB for the period 1998-2007.

Datasets	APHRO-DITE	APHRO-DITE-2	CPC	GPCC-daily	TMPA	GPCP-1DD
APHRODITE-2	0.920					
CPC	0.797	0.775				
GPCC-daily	0.819	0.836	0.816			
TMPA	0.760	0.762	0.687	0.712		
GPCP-1DD	0.735	0.725	0.665	0.676	0.898	
ERA5	0.888	0.903	0.743	0.810	0.741	0.727
ERA-Interim	0.854	0.870	0.722	0.777	0.733	0.727
JRA	0.843	0.860	0.677	0.759	0.702	0.697
MERRA2	0.846	0.862	0.714	0.778	0.708	0.699
MERRA1	0.834	0.849	0.683	0.760	0.698	0.688
CFSR	0.795	0.820	0.640	0.740	0.641	0.625
NCEP2	0.706	0.731	0.552	0.661	0.577	0.545
NCEP1	0.760	0.769	0.606	0.687	0.619	0.598
20CR	0.596	0.635	0.512	0.567	0.481	0.478
ERA20C	0.754	0.746	0.646	0.691	0.644	0.643

correlations between the two types of datasets is not affected by common data or method, and is rather a measure of their quality, which helps identifying the best datasets in each group.

We continue the comparison of the observational datasets using reanalyses as a reference (comparison along the rows of Table 2.6). APHRODITE-2 has systematically higher correlation with the reference, regardless of the reanalysis used, than the other observational datasets. It is followed by APHRODITE. Both have significantly higher correlation than GPCC-daily, in third position. By contrast, CPC has systematically a lower correlation than GPCC-daily. Interpreting these results in terms of quality, we attribute the lower performance of CPC and GPCC-daily to the much lower number of observational inputs than in APHRODITE and APHRODITE-2 (Table 2.2). Despite a slightly higher number of measurements, CPC performs worse than GPCC-daily, likely due to issues on the quality of those measurements, discussed in Section 2.3.1.1. Regarding satellite-based datasets, TMPA systematically outperforms GPCP-1DD, but the two, along with CPC, have the lowest correlations with the reanalyses. That is, satellite measurements seem to degrade the signal from rain gauge measurements.

We can also compare the reanalyses quality using observational datasets as a reference (along the columns of Table 2.6). ERA5 has systematically higher correlations with the observations. However, this reanalysis assimilates rain gauge measurements, such that it

Table 2.7 Same as Table 2.6 for the LIB

Datasets	APHRO-DITE	APHRO-DITE-2	CPC	GPCC-daily	TMPA	GPCP-1DD
APHRODITE-2	0.887					
CPC	0.838	0.825				
GPCC-daily	0.864	0.841	0.870			
TMPA	0.829	0.869	0.790	0.809		
GPCP-1DD	0.771	0.801	0.720	0.740	0.906	
ERA5	0.858	0.871	0.805	0.826	0.835	0.772
ERA-Interim	0.828	0.837	0.763	0.794	0.790	0.744
JRA	0.719	0.760	0.709	0.708	0.760	0.730
MERRA2	0.777	0.794	0.723	0.763	0.725	0.677
MERRA1	0.782	0.796	0.749	0.760	0.775	0.741
CFSR	0.700	0.690	0.626	0.657	0.672	0.618
NCEP2	0.601	0.632	0.572	0.618	0.576	0.523
NCEP1	0.635	0.643	0.605	0.623	0.596	0.545
20CR	0.442	0.400	0.350	0.393	0.345	0.308
ERA20C	0.655	0.712	0.643	0.663	0.678	0.673

is not completely independent from the observational datasets. It is certainly a sign of good quality that the reanalysis output resembles the observations, but the reanalysis data could also include some of the observation errors. ERA-Interim has the second highest correlations, and is the best performing reanalysis among those that do not assimilate precipitation observations. It is closely followed by MERRA2, while CFSR has poorer results among the latest generation of reanalyses. Interestingly for NCEP's reanalyses, the first version outperforms the second version. The two twentieth century reanalyses also show interesting behaviour: while 20CR has the lowest correlations with the observations, ERA20C performance is between CFSR and NCEP1, despite only assimilating surface observations. This behaviour clearly shows the progress made in reanalysis processing (e.g. in atmospheric modelling and data assimilation) over the last decades.

The same correlation analysis is performed for the LIB (Table 2.7). The results are quite similar, but we also note some interesting differences. The correlations between the observations are all higher for this study area. In this flat area, precipitation is less heterogeneous, and observations are more representative of their surrounding (i.e. larger spatial representativeness). In contrast, the reanalyses have lower correlations with observations than for the UIB. The LIB only receives precipitation during the summer monsoon, which is less well represented in models than the winter precipitation in the UIB (see following Section on

seasonality). More in detail, APHRODITE-2 and APHRODITE still perform best among the observational datasets, but the four other datasets rank in a different order: satellite products are possibly better in that flatter area. For the reanalyses, we noticed that MERRA2 does not outperform MERRA1. It echoes the large change in precipitation amount between the two discussed above (Table 2.4), and, similarly, could be related to the integration of CPC in MERRA2. Indeed, Table 2.7 suggests that CPC does not perform as well as the other observational datasets in terms of variability, and, indeed, surface moisture content variability was not improved from MERRA1 to MERRA2 in the area (Figure 1 in Reichle et al., 2017). As for ERA5 and ERA-Interim, they remains the two reanalysis datasets with the highest correlation with the observations.

2.3.3.3 Influence of the seasonality

Figure 2.4 presents the seasonality, for the UIB, of the correlations between the reanalyses and APHRODITE-2. This reference is chosen because of its higher correlation with the reanalyses, but the other rain gauge-based datasets give a similar seasonality. The Figure shows that the reanalyses are altogether more similar to APHRODITE-2 during the winter season than during summer. From December to April, all reanalysis products have a similarly high correlation with the observational dataset (>0.9), except for the two century reanalyses, and to a lesser extent the older NCEP reanalyses. From May onward, all correlations drop to various degrees. Both NCEP reanalyses drop the most, followed by CFSR. ERA5 shows the highest correlations, just above ERA-Interim, JRA, MERRA1, and MERRA2. For the century reanalyses, 20CR drops to very low values (<0.5 and even <0.2 in September and October), while ERA-20C remains at acceptable levels, around CFSR. Accordingly, we have very high confidence in the capability of most reanalyses to represent the daily variability in winter. In summer, the confidence is more dependent on the reanalysis, and overall lower than in winter. However, it is unclear if the seasonality of the correlation between APHRODITE-2 and the best reanalyses (ERA5, ERA-Interim, JRA, MERRA1, MERRA2) is due to a changing ability of the reanalyses or of APHRODITE-2. The seasonality for those reanalyses disappears when using TMPA as a reference, but mainly due to a drop in winter correlation, which rather suggests that satellite observations are not suited for that season (not shown). The analysis of the seasonality is less interesting in the LIB, since it is mainly dominated by the monsoon. The results resemble what was just discussed for summer in the UIB.

2.3.3.4 Trends

We also looked at possible trends in the representation of the daily variability, due to a change in the type, quantity, or quality of input data in each dataset. We computed the time series of correlations between observations and reanalyses using a two year moving window. However, the Pearson correlation we used so far is also known to be sensitive to extreme values. This leads to jumps in the correlation when an extreme value (unusually large precipitation event) passes in the moving window and is well represented. In order to have a clearer signal, without jumps, we used instead the Spearman correlation. This coefficient is based on the rank rather than on the absolute value of each observation and is therefore not sensitive to extreme values. We checked that most of the results presented above are valid with the Spearman correlation as well.

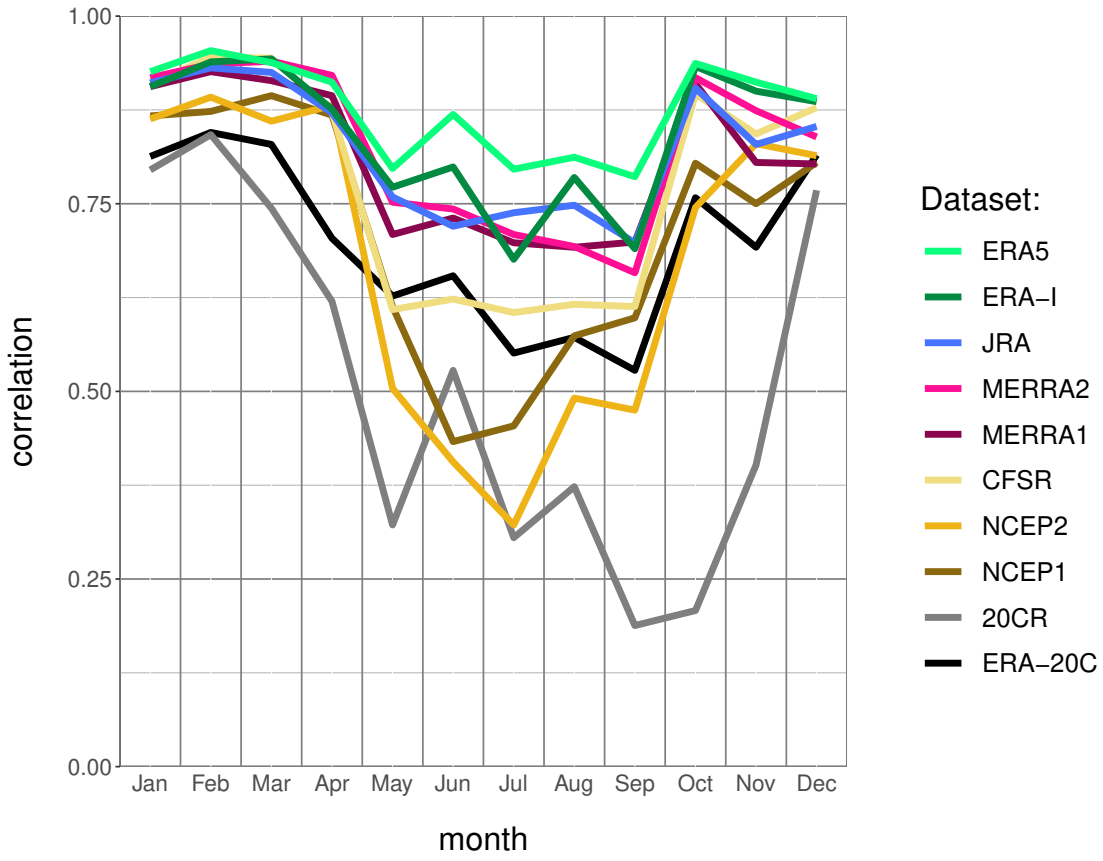


Fig. 2.4 Daily correlation, per month, between APHRODITE-2 and each reanalysis, in the UIB. The period considered is 1998-2007.

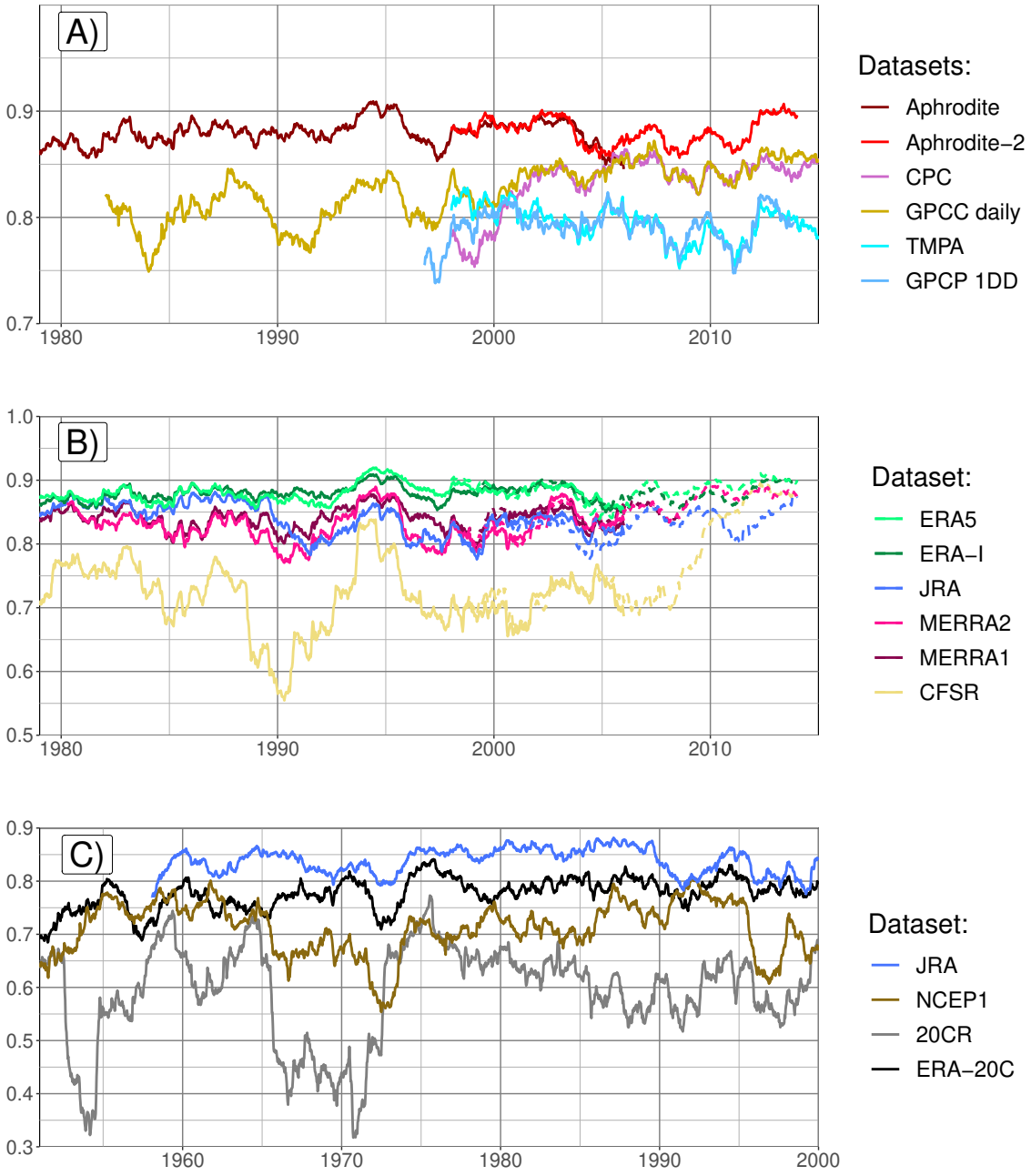


Fig. 2.5 Daily correlation using the Spearman formula, on a running two-year window, between a reference and different datasets, for the UIB. The years on the x-axis is the start of the two-year window. In A) observational datasets are tested against ERA-Interim. Figure B) shows the correlation between a selection of reanalysis and APHRODITE over the period 1979-2005 (plain line) and APHRODITE-2 over the period 1998-2013 (dotted line). Finally, C) presents the reanalyses covering the second half of the 20th century, with APHRODITE as reference

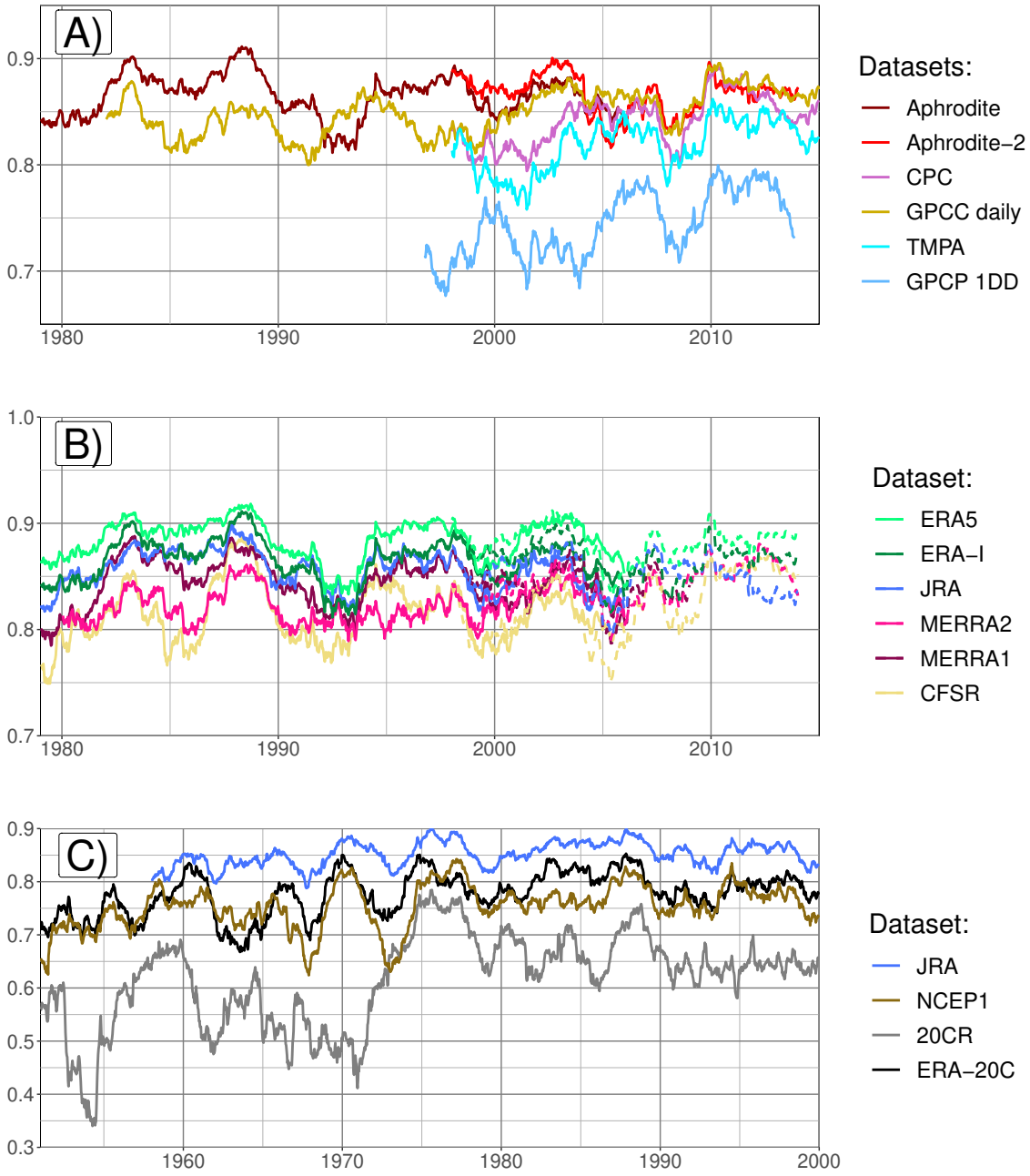


Fig. 2.6 Same as Figure 2.5 but for the LIB

In Figure 2.5-A, we compare the observational dataset using ERA-Interim as a reference, for the UIB. We first notice that APHRODITE and APHRODITE-2 always have significantly higher correlation scores than the others, except around 2004-2006, and relatively stable values between 0.85 and 0.9. The quality of those two datasets found over the period 1998-2007 can therefore be extended to the whole period 1979-present. GPCC-daily exhibits stronger variability during the first 20-years, but then its score increases and stabilises around 0.85. This behaviour is likely due to an increase of the number of observations that are between 5 and 10 before 2000, but above 15 after 2005. CPC is in general very close to GPCC-daily, except around the year 2000, which explains the differences between the two datasets over the period 1998-2007 previously investigated. The two satellite products TMPA and GPCP-1DD are very similar to each other, relatively stable, but at a lower level than the rain gauge-based datasets.

We now investigate in Figure 2.5-B the quality of the most recent reanalysis using as reference APHRODITE (plain line) and APHRODITE-2 (dotted line). These references are justified by the stability of their good results discussed above. They give similar results over their common period, which helps when analysing the whole time period. ERA5 and ERA-Interim are the two most stable reanalyses and have the highest correlations. JRA is also one of the best reanalysis datasets in the 1980's, but its correlation drops by about 0.05 compared to ERA5 after 1990 and never recovers. MERRA1 and 2 exhibit similar variability to each other, but the first version often has better results than the latter. CFSR is the most problematic reanalysis with the strongest variability and much lower correlation. However, it shows much better results at the end of the time period, with the release of its second version.

Lastly, over the second half of the twentieth century, the large change in number and type of observations assimilated could impact the quality of the reanalysis and is therefore investigated in Figure 2.5-C. However, no trend can be found. Correlations between JRA and APHRODITE remain mostly between 0.8 and 0.85. ERA-20C is also fairly stable over time, generally above NCEP1. 20CR, by contrast, exhibits a much higher variability with correlation dropping as low as 0.4 at times, and sometimes reaching NCEP1.

There are some differences in the results for the LIB as shown in Figure 2.6. First, for the observation, CPC and GPCC-daily reach the quality of APHRODITE-2 around 2005, despite including half the number of observations (Figure 2.6-A). Certainly, after 2005, the more homogeneous coverage of observations in CPC and GPCC-daily than in APHRODITE-2 counterbalances the reduction in number (Figure 2.1-D and E). Before 2005, the cause of the improvement of GPCC-daily can again be tracked to the increase in observations included, while the rise in quality of CPC remains of uncertain origin, since the number and location of observations are constant. TMPA shows correlation very close to CPC, with a similar

unexplained rise between 2000 and 2005, almost reaching the quality of the rain gauge-based datasets. GPCP-1DD has lower scores than TMPA, but also sees a rising trend during the two decades it covers. Comparing the differences between the reanalyses (Figure 2.6-B), we found much smaller differences than when using the Pearson correlation (Table 2.7), which suggest that the difference in quality resides in the representation of the extreme events. No clear change can be observed during the period 1979-2015, however.

2.3.4 Monthly, seasonal, and inter-annual variability

A good representation of daily precipitation variability does not ensure a good representation of monthly or longer period variability. Moreover, all the observational datasets selected for this study can be analysed at a monthly time scale. In Figure 2.7, we present the trend in monthly correlation between a reference and each type of dataset for the UIB. The correlation is calculated with the Pearson formula and over a ten-year moving window. It uses the monthly anomaly of precipitation, relative to a monthly mean computed over the same ten-year moving window. The reference to validate the observational datasets is ERA-Interim (A), and to validate the reanalyses GPCC-monthly (B). Those two datasets present a more stable quality and good correlations as we demonstrate below. They also cover the whole period 1979 to the present. However, we checked the main results with other references to validate them.

The best observational dataset for representing monthly variability for the UIB is APHRODITE (Figure 2.7-A). By contrast to the daily variability analysis, APHRODITE-2 has a significantly lower correlation with ERA-Interim on the common period with APHRODITE (1998-2007) and the correlation continues to drop after it. The difference in correlation between the two datasets is quite dependent on the reference, but all show the subsequent decrease. By contrast, CPC starts with the lowest correlation, but the correlation rises in the last decade at the level of the other datasets. CMAP, based on CPC also presents lower correlation, but is more variable, and it depicts a similar rise around the year 2000. All the other datasets are very close to each other.

Still for monthly variability, the closest reanalysis to the observations is ERA5 (Figure 2.7-B), except when using CPC and CMAP as reference: then, MERRA2 has higher correlation at times, likely due the use of CPC data in both CMAP and MERRA2. Several datasets show a decrease in correlation during the 1990s: JRA, has a drop more pronounced than what is observed for the daily variability, and a drop appears for NCEP1, NCEP2 and ERA-20C. 20CR has the lowest correlation, while MERRA2, MERRA1, and ERA-Interim are quite similar, with correlation just below ERA5. CFSR also has relatively high values, but exhibits a decreasing trend, especially in the last 10 years, which is even more pronounced when

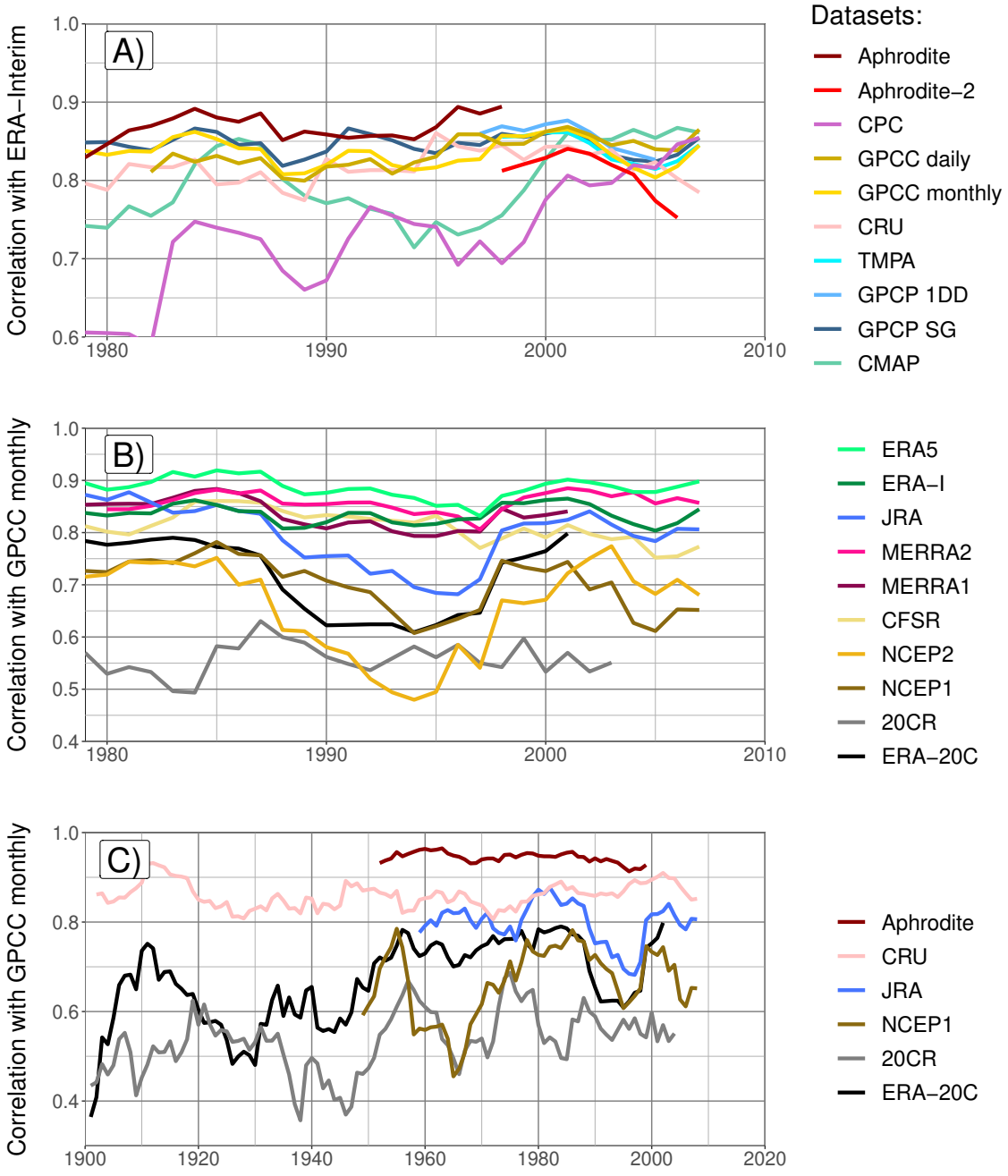


Fig. 2.7 Correlation of monthly anomaly on a running ten-year window for the UIB. The monthly mean needed for the anomaly is computed relatively to the ten-year window. The years on the x-axis is the start of the ten-year window. Similarly as in Figure 5, a set of datasets is tested against a reference. In A) observational datasets are tested against ERA-Interim. B) shows the correlation between the reanalysis and GPCC-monthly. Lastly, C) presents the longest datasets, except GPCC-monthly which is used as reference.

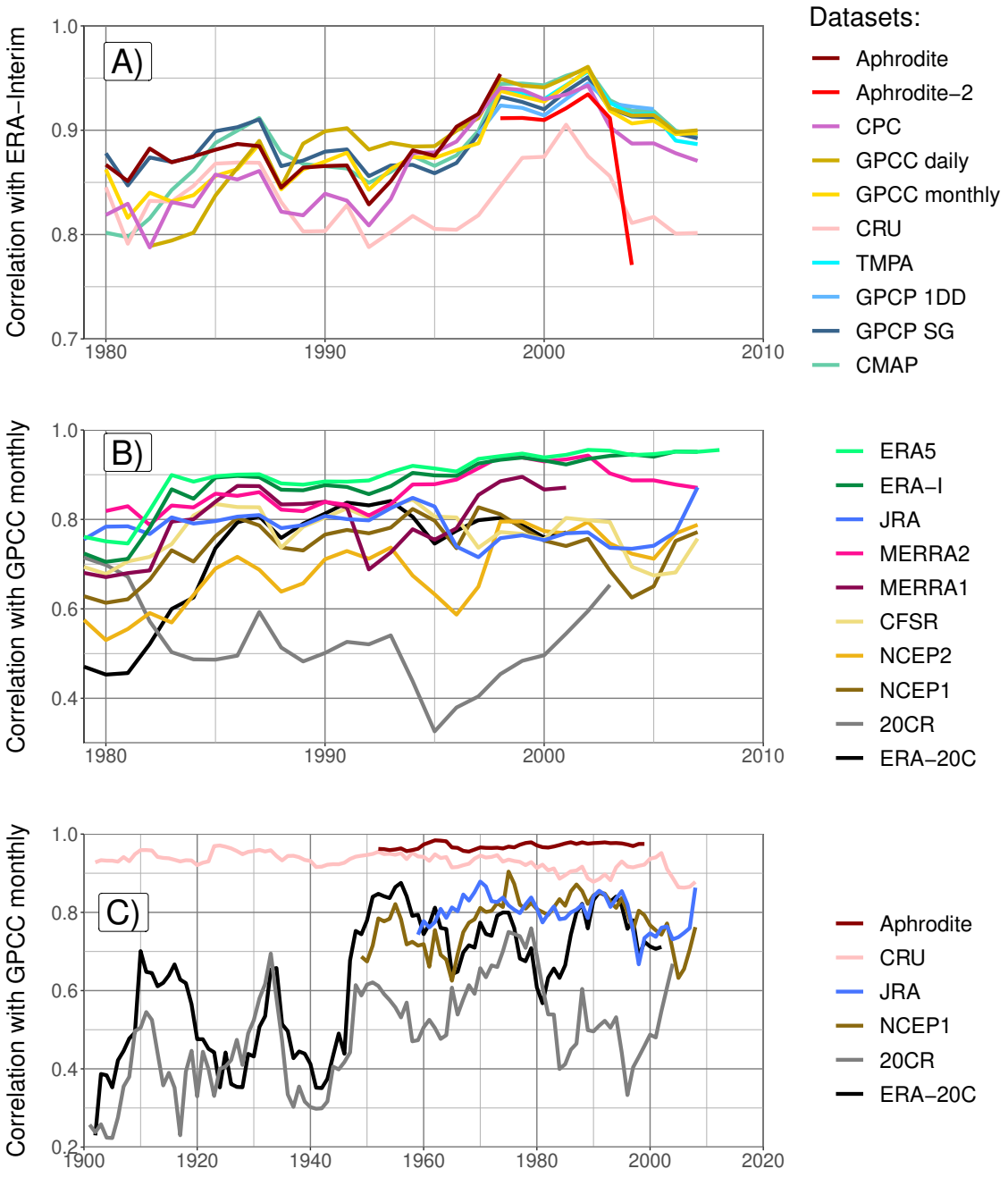


Fig. 2.8 Same as Figure 2.7 but for the LIB

testing with the other observational datasets. It is possible that version 2 of CFSR gives better results, but it has not been running long enough to evaluate the monthly variability over a 10-year period. Instead, the correlations in Figure 2.7-B include both versions toward the end of the time period, which could add discrepancies when computing the monthly mean anomaly.

We also tested the datasets with the longest time coverage against GPCC-monthly (Figure 2.7-C). We found relatively stable correlations with APHRODITE and CRU during the twentieth century: the time series do not diverge, despite the lowering number of observations. However, since the datasets are not independent, we cannot say that the quality of those datasets remains constant. The reanalyses present fluctuating correlations with the reference. ERA-20C has lower correlations in the first half of the century, which could be due to a lowering confidence in either the reference or the reanalysis. However, ERA-20C correlations get closer to 20-CR during that period, which suggests that the variation in the reanalysis quality is the most important factor.

The LIB shows somewhat different results in terms of monthly variability (Figure 2.8). For the observations, APHRODITE does not have the highest correlations, as it is bypassed by GPCP-SG during the 1980s. After 2000, all datasets perform very similarly with two exceptions: CRU, which always has lower correlations, and APHRODITE-2 whose correlations drop during the last two years. For the reanalysis, ERA5 still has the highest correlation but is joined by ERA-Interim just before the year 2000. MERRA2 does not show specifically higher correlation with CPC, as it does for the UIB, except for the two first years, where

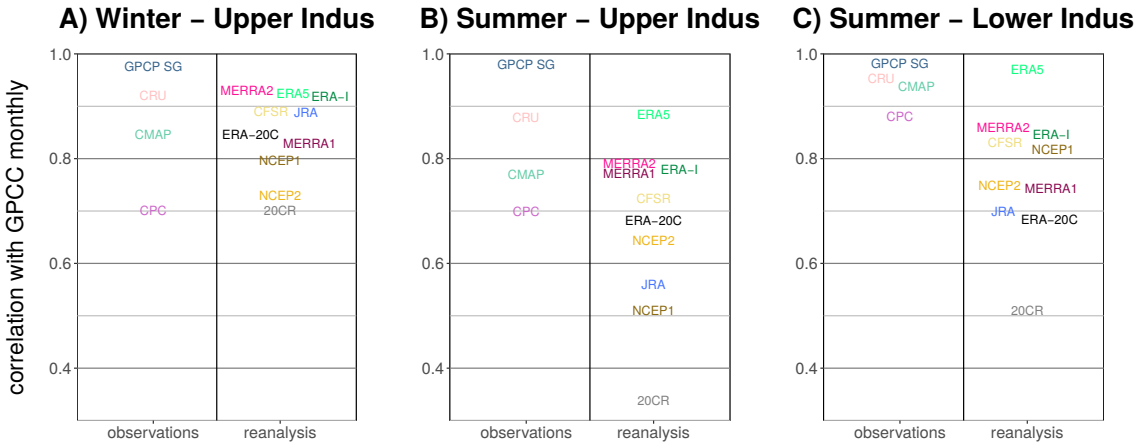


Fig. 2.9 Inter-annual correlation on the period 1981-2010 between GPCC-monthly and the other datasets covering that period. The correlations are computed for specific seasons and domains. We split the result by type of dataset (observation and renalysis)

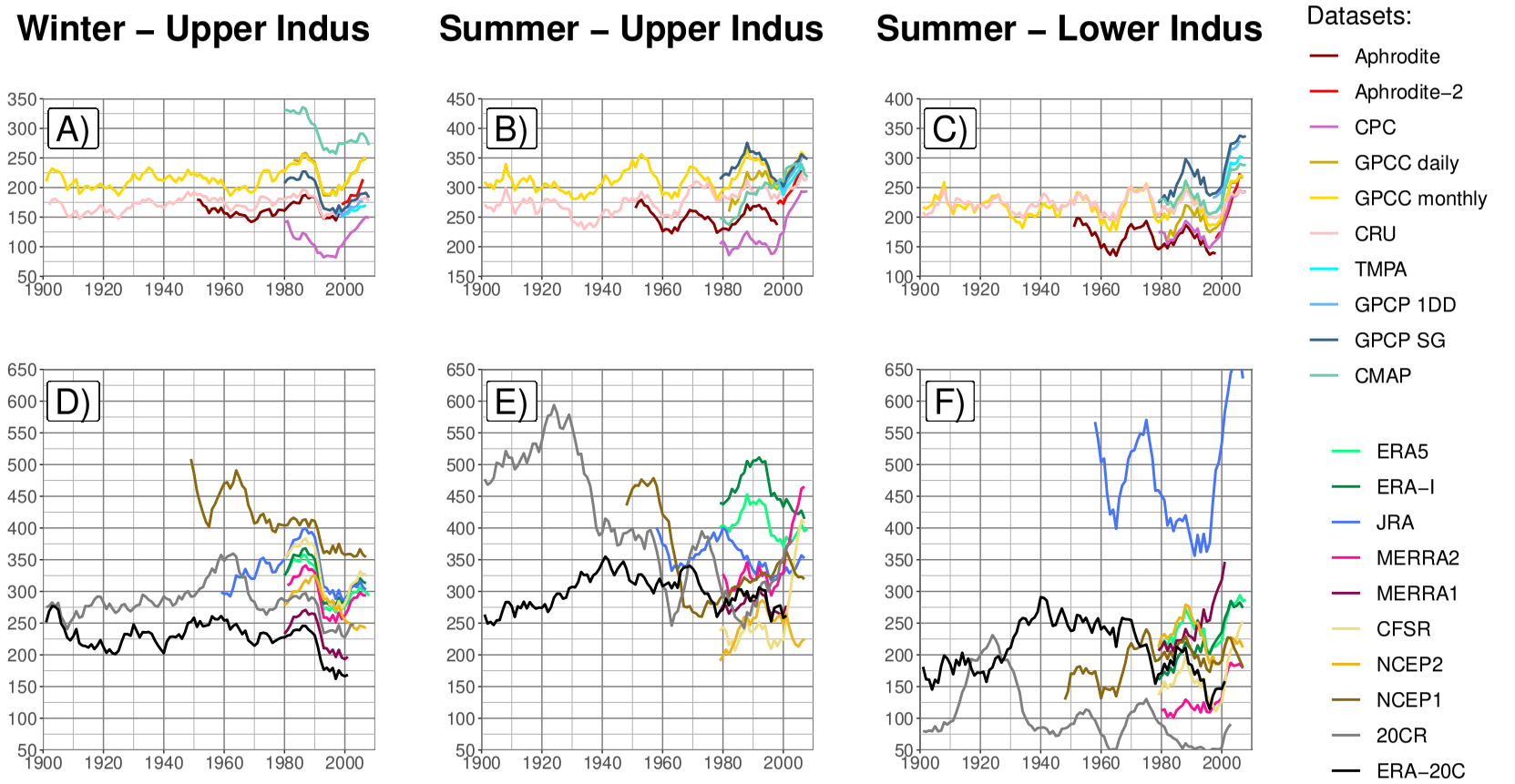


Fig. 2.10 Decadal variability of precipitation using a 10-year running mean for different seasons and domains (Winter in the UIB: A and D; summer in the UIB: B and E; summer in the LIB: C and F) and the different datasets (Observational datasets: A, B and C; Renalysis datasets: D, E and F)

CPC has the lowest values. It is possible that the smaller difference in quality between CPC and the other observational datasets is not important enough to influence MERRA2's quality significantly. Lastly, for the century-long datasets, correlations between CRU and GPCC-monthly show a decreasing trend, that could be related to an increasing difference in the observations included in each dataset. By contrast, ERA-20C correlation are as low as 20CR before 1950.

In Figure 2.9, we compare the inter-annual variability of GPCC-monthly to the reanalyses over the period 1981-2010 and to the other observational datasets covering that period. In Figure 2.10, we look at the 10-year moving mean for each of these datasets. Note that the years we mention in the text correspond to the start of that 10-year window. The results are split by season and study area. GPCP inter-annual variability is almost identical to that of GPCC-monthly, due to the inclusion of GPCC-monthly data (Figure 2.9). By contrast, CPC has a much lower correlation with GPCC-monthly, especially in the UIB. This agrees with the lower capabilities found for the daily and monthly variability of CPC. Moreover, CPC is the most dissimilar observational dataset for the decadal variability, particularly for the UIB, along with CMAP and APHRODITE-2 (Figure 2.10). In contrast, the other datasets show a very similar behaviour.

The reanalyses in winter have a decadal variability similar to the observation for the period 1980-2010 (Figure 2.10-D). Moreover, the most recent reanalyses tend to converge towards the same amount of precipitation after 2000. By contrast, the reanalyses that run before 1980 do not represent the decadal variability depicted by the observations. For summer in both study areas, none but ERA-5 represents the decadal variability observed. For example, in the UIB during summer, the precipitation amount increases after 2000 in the observations (Figure 2.10-B). While MERRA2 and CFSR show an increase of precipitation 2 or 3 times more important, ERA-interim and NCEP1 and 2 show instead a decrease (Figure 2.10)-E. Interestingly, while the observations show similar decadal variability for summer between the UIB and the LIB, this is not the case for the reanalyses, except maybe for the twentieth century reanalyses, and ERA-5. Notably, ERA5 has an inter-annual correlation with GPCC-monthly that is higher than the correlation between GPCC and CRU for all three panel in (Figure 2.9), suggesting it is at least as able as observational datasets.

2.4 Conclusion

In this study, we have compared a large number of precipitation datasets of different types across two distinct zones of the Indus watershed: six datasets are based only on rain gauges, four are derived from satellite observations, and ten from reanalysis. We have shown that the

number and diversity of the datasets help to identify and quantify the limitations and abilities of each of them, which in turn enables a better estimation of the uncertainties.

We have compared the datasets on the basis of the annual mean precipitation, the seasonal cycle, as well as the variability over time scales from one day to 10 years. We have relied on the literature to evaluate the different sources of uncertainty and have interpreted the mean differences between datasets in terms of their quality. We have suggested that the similarities in variability can directly be interpreted in terms of quality, especially when comparing datasets with no common methods or data source. Most reanalyses do not assimilate precipitation observations, which makes it possible to cross-validate between observational and reanalysis data based on variability. Regardless of the observational datasets used as a reference, we have found that some reanalyses have significantly higher correlation with that reference than other reanalyses, which we have interpreted as a sign of good quality. Conversely, when using a reanalysis as a reference, some observational datasets have significantly higher correlation than others. The use of reanalyses to validate observational datasets is justified by the quality of reanalysis products demonstrated in this study. Specifically, at the scale of the Indus Basin, and for the daily variability, the same level of similarity between the reanalyses and observations is also seen between the observational datasets themselves.

We have used the Pearson correlation to compare the datasets, although this has some limitations. For example, it is affected by extreme values, that is, in our context, unusually large precipitation events. These lead to difficulties in interpreting trends and we preferred the Spearman formula in this context (cf. Figures 2.5 and 2.6). By contrast, the Pearson correlation is less affected by the difficulties in representing the lowest precipitation rates, although these rates can explain some of the biases.

One of our findings concerns the important uncertainty in fine scale spatial patterns of precipitation, particularly in the UIB, where precipitation is the most heterogeneous. Important discrepancies remain between datasets, which explain part of the differences in mean precipitation. This issue needs to be tackled in observational datasets by including more measurements and by updating the climatology used in the interpolation methods. In reanalysis products, higher resolution and better modelling of small scale processes are likely needed to improve confidence in the spatial pattern of precipitation. In this study, we have deliberately selected two large study areas, which has increased the confidence in the datasets. Area-wide correlation particularly improves the significance of the variability analysis, compared to a point-wise correlation.

We have also found that the quality of the datasets depends on the season. Rain gauge measurements suffer from important underestimations in winter for the UIB. Most satellite-

derived datasets even further amplify this bias. By contrast, reanalyses perform best during winter. Particularly, the most recent reanalyses produce a very similar amount of winter precipitation and its variability is similar to the observations at all timescales. We have suggested that their amount of precipitation is closer to reality than the observations, although some overestimations are possible, due to, for example, misrepresentation of the lowest precipitation rates. Summer precipitation, in both study areas, is much more uncertain in the reanalyses in total amount, seasonality, and variability. In contrast, satellite observations perform better in summer than in winter and seem to bring additional information to rain gauge measurements.

As mentioned above, rain gauge-based datasets underestimate precipitation. Only GPCC products use a correction factor to account for measurement underestimation, but this factor is still too small. We emphasise the need to correct directly the measured values before interpolation to a grid dataset, using, for example, methods similar to those developed by Dahri et al. (2018).

More specifically, APHRODITE is the best observational dataset for daily and monthly variability, thanks to a large number of observations in the whole basin. However, it also exhibits drier conditions than most of the other datasets, which is partially caused by the interpolation method it uses and possibly by a lower quality of the data. Surprisingly, APHRODITE-2 is not as good, especially for the longer term variability, as it removes some observations in areas with an already lower density of measurements. CPC is the least reliable observational dataset, particularly for the UIB, with a large dry bias compared to GPCC-monthly, the lowest correlation scores at all time scales, and an error on the dates before 1998. However, its quality significantly improves after 2005, which, we suspect, is due to a change in the quality of the data source. GPCC-monthly is one of the most reliable datasets both in terms of amount and variability. GPCC-daily relies on GPCC-monthly for its monthly mean. The very low number of daily measurements included in the early part of the covered period limits its quality, but this quickly improves as more observations are included.

Satellite-based datasets are very dependent on the quality of the rain-gauge product they integrate. The added-value of satellite observations remains limited at the basin scale. The signal is degraded during winter for the UIB, while better results in the LIB suggest slightly wetter conditions than the rain gauge-based datasets. Importantly, the quality of satellite-based datasets resides in their near real time availability as well as their higher temporal and spatial resolution than rain gauge based datasets.

The quality of reanalysis datasets has clearly improved since the first datasets were released. ERA5 is the latest reanalysis and clearly stands out as the one representing best the observations, in terms of amount, seasonality, and variability at all time scales investigated.

Remarkably, it is the only reanalysis representing the decadal variability of the summer precipitation for both study areas as it is seen in the observations. Furthermore, for the daily to inter-annual variability, the best performing observational dataset has often a better level of similarity with ERA5 than with other observational datasets. Some of these qualities can be derived from its high resolution, which allows the representation of interesting fine scale features, as well as the assimilation of precipitation measurements.

After ERA5, ERA-Interim, MERRA1, and MERRA2 have relatively similar performance. Reichle et al. (2017) showed that the soil moisture content was not improved over South Asia from MERRA1 to MERRA2, neither in terms of variability nor biases, despite the use of CPC to correct the precipitation input to the land surface model of MERRA2. Given the difficulties of CPC to represent precipitation in the Indus Basin, correcting the modelled precipitation with this dataset probably does not improve the signal. In this study, we were able to show that the correction with CPC feeds back locally on the modelled precipitation, particularly at the monthly scale for the UIB. We have also suggested that the dry bias of MERRA2 in the LIB, and the decrease score on the daily variability compared to MERRA1, is also due to that correction.

The confidence in JRA's precipitation in the UIB is generally high, but drops for the daily and monthly variability in the 1990's. By contrast, it represents overly wet conditions for the LIB. CFSR has problems reproducing the daily variability and the seasonality of the monsoon, especially in the UIB. This is probably improved by the latest version that started in April 2011. However, it would likely be better to treat the two versions separately as it seems the new version produces somewhat different statistics of precipitation. The twentieth century reanalyses, which includes only surface observations, are not as good as the others, especially in winter. However, while 20CR barely reproduces any of the variability depicted by the observation, ERA-20C has much better capabilities, close to NCEP1 and CFSR, especially during summer. Neither 20CR nor ERA-20C represent the decadal variability seen in the observation before 1980.

Finally, large uncertainties remain about precipitation in the UIB, but one should not treat all datasets equally. We have demonstrated that specific datasets represent the precipitation better, which helps to narrow down the uncertainty. Particularly, we have argued that precipitation from reanalyses and observational datasets can both be useful for cross-validation. They can also be used for quality monitoring. Daily correlation of precipitation for key areas can be performed between a series of datasets with near real time updates. Changes in correlation between one or several datasets would therefore highlight a change in quality that would need to be investigated.

ERA5 reanalysis will be used as the main reference for model evaluation in this thesis, due to two qualities. First, at the scale of the Indus Basin, it produces timeseries of precipitation as reliable as rain gauge-based datasets. Second, it produces 3D atmospheric fields that can also be compared to model outputs in order to understand the origin of precipitation differences (Chapters 3, 4, and 5). For longer timescales, GPCC-monthly will also be considered (Section 5.4.1).

Chapter 3

Contribution of cross-barrier moisture transport to precipitation in the UIB

Preface

The content of this chapter, both in terms of methodology and results, have been published in Baudouin et al. (2020a). Co-authors C.A. Petrie and M. Herzog provided guidance and reviewed the paper. Only small modifications have been performed on the introduction and conclusion compared to the published version, so the chapter can fit in the thesis.

3.1 Introduction

Chapter 2 established that ERA5 is one of the observational datasets that represents precipitation seasonality and variability in the Indus River Basin the most accurately. As a reanalysis, it also provides 3D atmospheric fields. These fields are used in this chapter to better understand the processes at the origin of precipitation variability and seasonality, so it can be established whether climate models can represent these processes. To reduce the scope of the analysis, only precipitation in the Upper Indus River Basin (UIB) is considered.

The upper Indus River Basin is a key water storage zone on which the population of Pakistan and north-west India have relied throughout the Holocene for water supply and agriculture, in an otherwise semi-arid or arid basin (Baudouin et al., 2020b; Petrie et al., 2017; Yu et al., 2013). The Indus watershed is dominated by a large flat plain and major mountain ranges including the Hindu Kush, the Karakoram, and the Himalayas that lie to its North. The upper Indus River Basin (UIB) is mostly characterised by this mountainous terrain.

Precipitation in the UIB mainly falls during two wet seasons which have different synoptic characteristics. In winter, from December to March, extra-tropical Western Disturbances bring rain, and snow at altitude, over most of the mountains and nearby plains (Dimri et al., 2015). In summer, mostly in July and August, the South Asian monsoon extends to the Indus River basin and produces large quantities of precipitation near the foothills of the Himalayas during intense episodes of deep convection (Barros et al., 2004; Houze et al., 2007).

Previous studies of precipitation in the Indus or upper Indus River Basin have noted the importance of the topography to trigger or enhance precipitation during both wet seasons and explained the differences between the mountainous areas and the drier lowland (e.g. Baudouin et al., 2020b; Dahri et al., 2018; Immerzeel et al., 2015; Iqbal and Athar, 2018; Ménégoz et al., 2013; Palazzi et al., 2013). These studies have stressed the uncertainties in precipitation amount and finer scale patterns in the mountainous areas due to the scarcity, heterogeneity and biases in precipitation observations. Numerical experiments have also been performed for case studies to better understand the processes involved in the generation of precipitation (Barros et al., 2004; Dimri and Chevuturi, 2014; Dimri, 2004; Medina et al., 2010; Muhammad Tahir et al., 2015, among others). Unsurprisingly, these authors have found that the topography is responsible for the increase of precipitation along the foothills. They especially show the importance of wind and up-slope moisture convergence, associated with convective instability, in the production of precipitation during both wet seasons. However, a systematic analysis of synoptic scale processes causing the dependency between precipitation and topography at the basin scale has not been performed so far.

Houze (2012) has reviewed the processes related to precipitation generation in mountainous areas. Most processes involve cross-barrier flow, when a mountain barrier dynamically forces the air to rise up to saturation if enough moisture is present (e.g. moisture convergence). Depending on the stability of the flow, the effect of mountains differs. In cases of high static stability or weak motion, the cross-barrier flow can be blocked and deflected by the mountain barrier, limiting convergence on the foothills. In contrast, if the flow is stronger, or the atmosphere has lower static stability, up-slope convergence is maximised. Convective instability can also provide the energy that sustains vertical velocities. This energy is released either by differential diurnal heating or by a cross-barrier flow. In the latter case, kinetic energy from the cross-barrier flow and convective potential energy combine to strengthen the vertical velocities. These strong interactions between instability and cross-barrier flow produce complex patterns in precipitation variability.

In this study, we investigate the processes that control precipitation and their relationship to the topography. We only consider the precipitation average over the UIB to avoid dealing with finer scale heterogeneity and associated uncertainty. Specifically, we quantify the

importance of the cross-barrier wind effect in the production of precipitation. The method is based on common statistical tools that are not often applied in atmospheric science. Subsequently, the analysis focuses on the similarities and differences between seasons, and the importance of the direction and the altitude of the cross-barrier flow. Lastly, we quantify the ability of climate models to represent these processes.

3.2 Data

3.2.1 Study area

The Upper Indus River Basin (UIB) is defined as in Baudouin et al. (2020b) as the part of the Indus River Basin north of a line between 68.75°E-33.5°N and 77.5°E-30°N, and is delimited by the black contour line in Figure 3.1. This study area includes the most mountainous part of the watershed: part of the Hindu Kush and the Sulaiman ranges to the west, the Karakoram to the north, and the western part of the Himalayas to the north-east. It also extends to the plains to the south, in order to incorporate most of the maximum of precipitation falling along the foothills (Figure 3.1, see also Baudouin et al., 2020b). Figure 3.1 also shows the topography as represented in the ERA5 reanalysis dataset. The topography rises steeply north of the plain to reach a maximum of between 3000 and 8000 m above sea level. The orientation of the mountain ranges changes from west to east, forming a notch where the Hindu Kush and the Sulaiman range connect to the Karakoram (See Figure 3.1).

3.2.2 Datasets

This study requires estimates of both precipitation and moisture transport, defined as the product of wind and specific humidity. Gridded estimates of atmospheric variables, such as wind and moisture, are only provided by reanalyses. These atmospheric variables are then compared with precipitation data from the same reanalysis. Doing so ensures consistency between the variables and lowers the statistical threshold for significant results.

Reanalysis data from ERA5 (Hersbach et al., 2018) are used for this study. ERA5 produces the most reliable precipitation estimates for the study area, both in terms of total amount and variability (Baudouin et al., 2020b). The data were selected over the maximum period 1979-2018 (40 years) at a 3-hourly time step. Precipitation corresponds to the accumulation over this 3-hourly period, averaged over the study area. Instantaneous values of wind and specific humidity are considered at three different standard pressure levels: 850 hPa, 700 hPa, and 500 hPa, and at each grid point (at 0.5° horizontal resolution) within the area between 28 and 40°N, and between 65 and 85°E. ERA5 data provide extrapolated values on

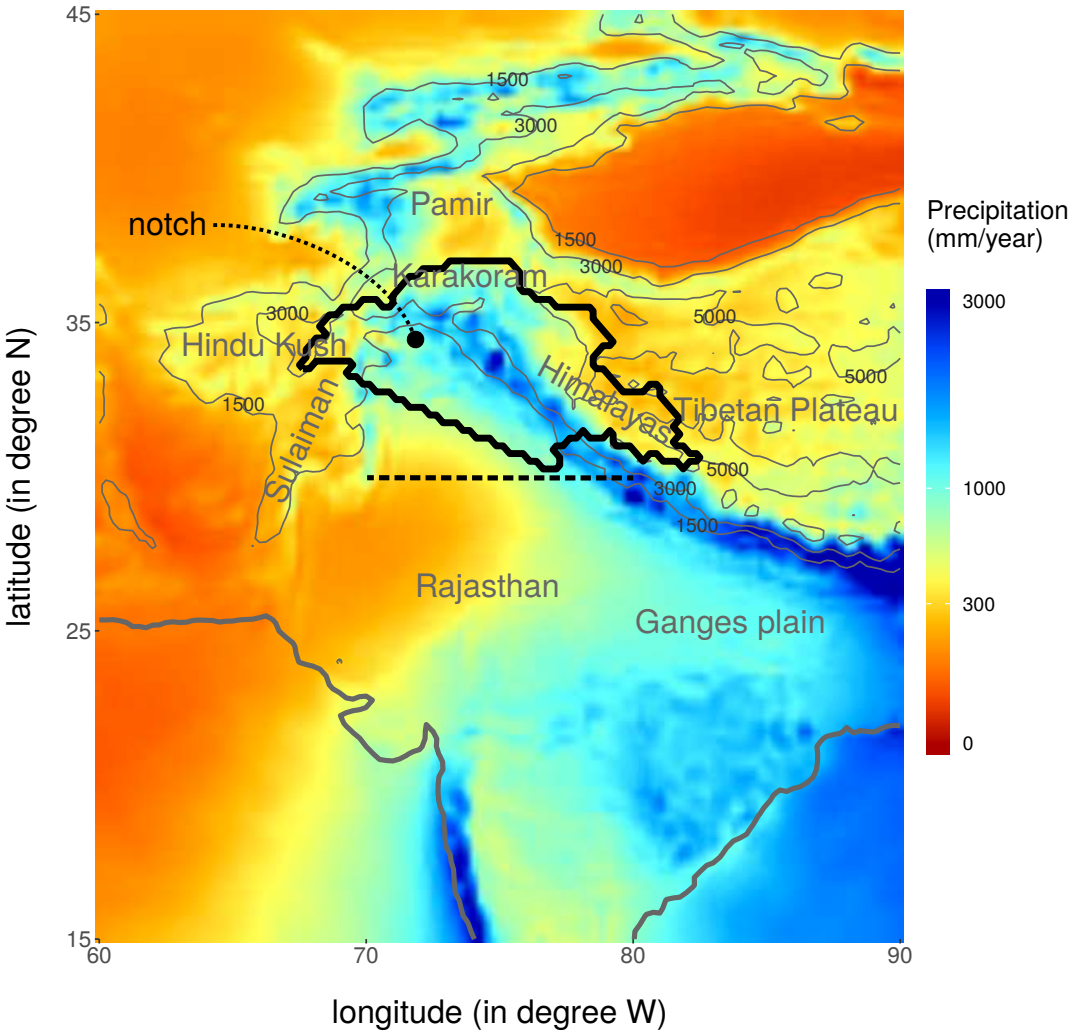


Fig. 3.1 Map of mean annual precipitation from ERA5 reanalysis dataset and for the period 1979-2018. The UIB is delimited by the black thick contour. The thinner contour lines represent the relief as modelled in the reanalysis (1500 m, 3000 m and 5000 m levels are indicated). The thicker grey line represents the coast line as in the reanalysis. The black doted mark the area where the mountain ranges form a notch. Finally, the dashed line is the cross-section used in Section 3.4.4.

pressure levels below the model surface. To ensure that for a specific grid point we use values mostly above the model surface (see contour lines in Figure 3.1), we deselect values for grid points where the model surface is above 1500 m for 850 hPa, 3000 m for 700 hPa, and 5000 m for 500 hPa. Note that the wind, and therefore the moisture transport, were separated into northward, westward, southward, and eastward directions. Higher temporal and spatial resolution are available for ERA5 but have not been used due to computing limitations and for a better consistency with climate model output.

Lastly, we use the output from climate models: the historical simulations produced for the Coupled Model Intercomparison Project phase 6 (CMIP6, Eyring et al., 2016). We use the data from the period 1980-2010 of this experiment. Note that in CMIP6, the instantaneous values of atmospheric variables are at a sub-daily resolution (6-hourly). As of August 2019, data for humidity and wind on pressure levels, and precipitation are available for the historical experiment from three climate models: GISS-E2-1-G, IPSL-CM6A-LR, MRI-ESM2-0 (Table 3.1).

Table 3.1 Available CMIP6 climate models, only the historical experiment is considered.

Model name	Short name	Variant	Nominal resolution	Citation
GISS-E2-1-G	GISS	r1i1p1f1	250 km	NASA/GISS (2018)
IPSL-CM6A-LR	IPLS	r1i1p1f1	250 km	Boucher et al. (2018)
MRI-ESM2-0	MRI	r1i1p1f1	100 km	Yukimoto et al. (2019)

3.3 Methods

3.3.1 Multilinear regression

The large-scale atmospheric circulation associated with precipitation events is often studied using a composite analysis, that is, by averaging the circulation over a large number of events (e.g. Hunt et al., 2018a; Midhuna et al., 2020; Vellore et al., 2016). However, this technique cannot determine which components of the circulation (location, altitude, and direction) are the most important in the generation of precipitation, nor can it quantify the strength of this link. Statistical methods help answering these questions. Here, we aim at predicting precipitation from moisture transport. We use a general linear model to fit precipitation (the predictand) with the time series of moisture transport at each grid point (the predictors) with the ordinary least square method. In mathematical terms, the model is defined as:

$$\mathbf{Y} = \widehat{\mathbf{Y}} + \mathbf{E}, \text{ where } \widehat{\mathbf{Y}} = \widehat{\beta}_0 + \sum_i (\widehat{\beta}_i \times \mathbf{X}_i) \tag{3.1}$$

In equation 3.1, the predictand $\mathbf{Y}$ is the time series of average precipitation over the UIB. With all 40 years of ERA5, the predictand includes $n = 116877$ observations, without distinguishing the seasons. The predictors $\mathbf{X}_i$ are the time series of southerly moisture transport at 700 hPa, defined as the product of wind and specific humidity. The time series are defined at each grid point within the box defined in Section 3.2.2 (see extent of the figure 3.2-A; note that points outside of the UIB are considered). There is a total of $p = 645$ predictors or grid points. The coefficient $\widehat{\beta}_i$ corresponds to the importance (or weight) of the i^{th} predictor to predict precipitation. The coefficient $\widehat{\beta}_0$ is the intercept of the regression, and is interpreted

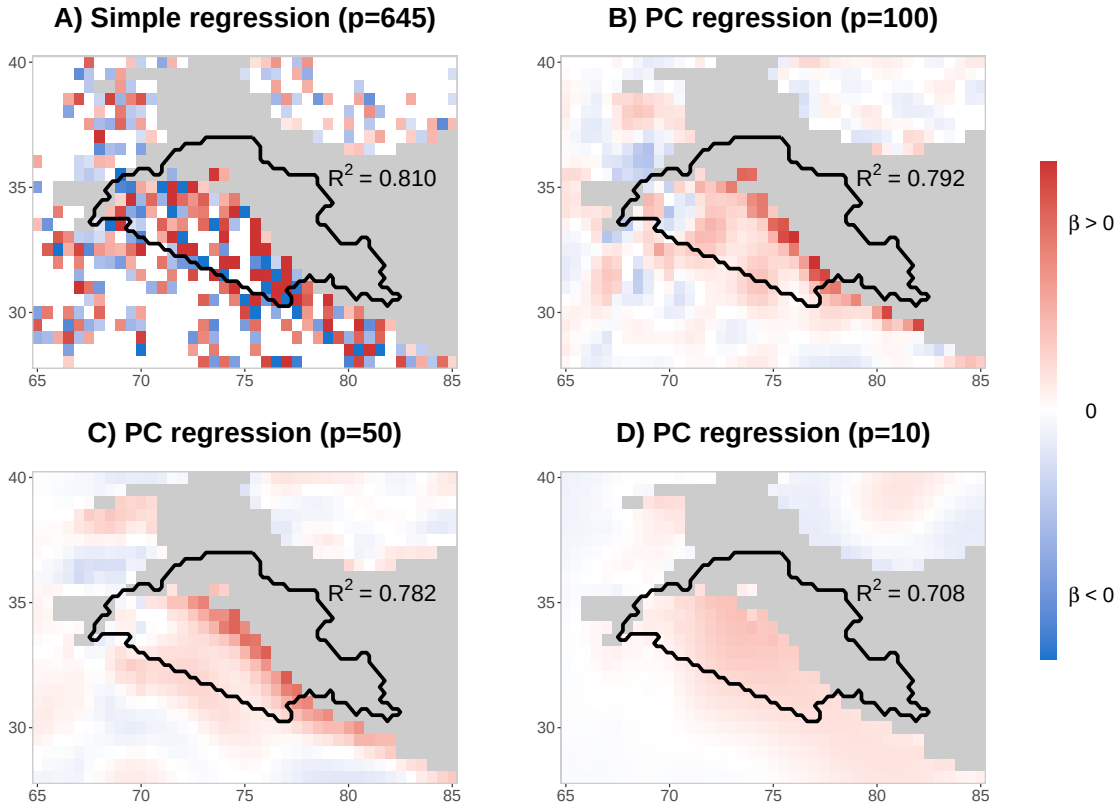


Fig. 3.2 Map of the coefficients given by different regressions between precipitation and southerly moisture transport at 700 hPa: in A, all the grid points are considered as predictor ($p = 645$); in B, C and D, the regression is made with the first $p = 100$, 50 and 10 PC (Principal Components) respectively. Non-significant values at the level 95% are in white, while grey areas represent excluded grid points (because of a surface above 3000 m). The validated coefficient of determination R^2 is given for each case.

as an offset. Note that these coefficients are independent of time. The prediction $\widehat{\mathbf{Y}}$ represents the part of precipitation, still as an average over the UIB, associated with moisture transport. Finally, the residual $\mathbf{E}$ (or model's error) is the part of the precipitation not related to moisture transport, and we associate this residual to small-scale processes such as convection.

The quality of the regression is defined using the coefficient of determination:

$$R^2 = 1 - SS(\mathbf{E})/SS(\mathbf{Y}) \tag{3.2}$$

where SS is the sum of square function (i.e. the variance multiplied by the degree of freedom). R^2 also equals the square of the Pearson's correlation coefficient between the predictand and the prediction. R^2 is validated on a sample of observations distinct from the one used to train the model: nine tenths of the observations are randomly selected to generate the training sample, while the remaining one tenth forms the validation sample. This selection is repeated 10 times to reduce the sampling uncertainty. The final R^2 equals the mean of R^2 for each instance of sampling.

The uncertainty of the estimate of the coefficients $\widehat{\beta_i}$ is evaluated using their standard deviation σ given by the equation:

$$\sigma(\widehat{\beta_i}) = \sqrt{\frac{SS(\mathbf{E})}{SS(\mathbf{X}_i) \times (1 - R^2_{\mathbf{X}_i}) \times df}} \tag{3.3}$$

where $R^2_{\mathbf{X}_i}$ is the coefficient of determination of the regression of $\mathbf{X}_i$ by all the other predictors $\mathbf{X}_{j \neq i}$, and df is the degree of freedom ($n - p - 1$, for n observations and p predictors). Finally, a Student's t-test is performed at the 95% confidence interval on the variance of the coefficients, to determine whether the coefficient is significant.

3.3.2 Dealing with multicollinearity

Figure 3.2-A presents the result of the regression between precipitation and southerly moisture transport at 700 hPa. With $R^2 = 0.81$, it has a high predictive skill. However, there is only a relatively limited number of significant coefficients despite the high number of observations, and the figure lacks a homogeneous pattern, which limits the interpretation. The cause of the noisy pattern is the spatial auto-correlation of the 2D field, also referred to as multicollinearity of the predictors in the context of a multilinear regression (see for example Saporta, 2006). The effect of multicollinearity, or the fact that a predictor can be predicted by the other predictors, is apparent in equation 3.3. There, the coefficient $R^2_{\mathbf{X}_i}$ measures the collinearity between the predictor $\mathbf{X}_i$ and the other predictors $\mathbf{X}_{j \neq i}$. In case of a high collinearity, $R^2_{\mathbf{X}_i}$ gets close to 1, which increases the standard deviation and thus the uncertainty on the coefficient

$\widehat{\beta}_i$. Eventually, the number of significant coefficients is reduced, as observed in Figure 3.2-A. Moreover, the ordinary least square method used to solve the regression equation involves a matrix inversion. Multicollinearity implies that the matrix is very close to rank deficiency. In that case, the inverted matrix becomes unstable and magnifies noise (Saporta, 2006), resulting in the patchy pattern seen in Figure 3.2-A.

A common way to reduce the impact of multicollinearity is the use of the Principal Component (PC) regression. It acts on the predictors themselves by reducing their number in the regression (in equation 3.1). The PC regression method consists first in a decomposition in principal components (or principal component analysis, PCA) of the predictors. During this process, the predictors are standardised. Then, a certain number of PCs, that need to be determined, are selected and used for the regression. The timeseries of each selected PC are then used as new predictors in the regression as shown in the following equation:

$$\widehat{\mathbf{Y}} = \widehat{\beta}_0 + \sum_{j \in selected\,PC} (\widehat{\beta}_j \times \mathbf{PC}_j) \text{ with } \mathbf{PC}_j = \sum_i (\alpha_{ij} \times \mathbf{X}_i) \tag{3.4}$$

$\mathbf{PC}_j$ is the timeseries of the j$^{\text{th}}$ PC, and is a linear combination of the timeseries of moisture transport at each location ($\mathbf{X}_i$). The coefficients α_{ij} form the rotation matrix from the PC analysis. The prediction $\widehat{\mathbf{Y}}$ can be written in function of $\mathbf{X}_i$:

$$\widehat{\mathbf{Y}} = \widehat{\beta}_0 + \sum_i (\sum_j \widehat{\beta}_j \alpha_{ij}) \times \mathbf{X}_i \tag{3.5}$$

It is clear from equation 3.5 that the prediction is still a function of the same timeseries $\mathbf{X}_i$ as in equation 3.1, despite having fewer coefficients $\widehat{\beta}_j$. The PC regression acts as a low pass filter: with fewer PCs selected, the result of the regression gets coarser (or less noisy). The objective is to select the number of PCs that allow the regression to resolve the extent of the phenomenon investigated. We are unable to find a satisfying statistic criterion for the selection: a reduction of the dimension fails to increase R^2, a stepwise regression based on the significant decrease of the variance of the residuals includes too many PCs, and a threshold based on the cumulative variance explained by the selected PCs is strongly dependent on the characteristic of the 2D field considered.

In our example, we apply the regression to the first 10, 50 and 100 PCs respectively. Figure 3.2-B, C and D show the resulting pattern by plotting each coefficient associated with $\mathbf{X}_i$ as shown in equation 3.5. The selection of the first 50 PCs (Figure 3.2-C) produces good result in terms of interpretability, with a pattern that overlaps with most of the study area.

3.3.3 Using several 2D fields

In the two previous sections, we used the southerly moisture transport at 700 hPa to predict precipitation. Now, we look at other predictors, especially different altitudes or directions of moisture transport, to ascertain whether they can further improve the prediction. The additional predictors, all 2D fields, are included in the regression along with the southerly moisture transport at 700 hPa. A decomposition in principal components (PCA) is still needed to address the multicollinearity between the predictors. However, we do not perform the decomposition of all 2D fields simultaneously since doing so would couple different fields and hinder the selection of fields that significantly contribute to the precipitation variability. In addition, we are interested in patterns of similar extent between each 2D field in order to compare them. A simultaneous decomposition would mask the distinct spatial details explained by individual fields. So, the PC decomposition is performed for each 2D field independently. The number of PCs for each field is defined so that each field is represented by a pattern of similar extent. A measure for the extent of a pattern is needed in order to define a consistent threshold across the different fields.

We consider a gridded map (or matrix) of n points filled with value α_i. When plotted, the values α_i form a pattern for which we want to measure the extent. The *Extent of Pattern* (EP) can be described by the root mean square of the differences of value α_i between all pairs of neighbouring grid points (i, j) of this map. We only consider neighbouring points of the same latitude or longitude (m pairs). This value is then normalised by dividing with the root mean square of the values α_i at all grid points. Hence, this normalised value represents the mean change between two neighbouring points in units of deviation from zero. The inverse of this value, multiplied by the mean distance (d) between neighbouring points, represents the mean distance needed to observe a change of one deviation to zero, and is used to define EP for a map of values α_i:

$$EP = d \times \sqrt{\frac{\sum_i \alpha_i^2 / n}{\sum_{(i,j)} (\alpha_i - \alpha_j)^2 / m}} \tag{3.6}$$

For example, for Figure 3.2-C, there are $m = 1167$ pairs of neighbouring points with a distance of 0.5°, the resulting *Extent of Pattern* is $EP = 1.13°$. In Figure 3.2-A, it reaches only $EP = 0.32°$, consistent with the much noisier pattern that is visible in this figure.

The effect of reducing the number of PCs selected in a PC regression can be re-interpreted with EP. In Figure 3.3, EP is computed for each PC for the southerly moisture transport at 700 hPa. The figure shows that the *Extent of Pattern* generally decreases for PCs that explain less variance (i.e. with a higher PCS number). Hence, limiting the regression to the PCs with a higher EP limits the noise in the regression pattern, as seen in Figure 3.2-B, C

and D. We determined in section 3.3.2 that a selection of the first 50 PCs gives satisfying results (Figure 3.2-C). The *Extent of Pattern* of the 50$^{\text{th}}$ PC is 0.73°, after smoothing. This value is used as a threshold to determine the number of PCs used for the other 2D fields. The third and fourth rows of Table 3.2 show the number of PCs selected for each field, as well as the cumulative variance explained for this selection. Differences in the cumulative variance explained can be observed between the different altitudes of moisture transport, which highlights the importance of not using explained variance if the goal is to have patterns of similar extent.

Equation 3.1 can be rewritten when considering several 2D fields, as follow:

$$\mathbf{Y} = \sum_{k} \widehat{\mathbf{Y}}_k + \widehat{\beta}_0 + \mathbf{E}, \text{ with } \widehat{\mathbf{Y}}_k = \sum_{j \in selected\,PC} (\widehat{\beta}_{jk} \times \mathbf{PC}_{jk}) \tag{3.7}$$

$\widehat{\mathbf{Y}}_k$ is defined as the "partial" prediction associated with the k^{th} 2D field, which corresponds here to the contribution of the moisture transport at a given altitude and direction to precipitation.

The next step, after performing the PCA and selecting the number of PCs for each 2D field, is to select the most important fields for the regression of precipitation. This selection is performed by using Pratt's Index (Pratt, 1987), which quantifies the relative importance of each predictor. Following the notation of equation 3.1, Pratt's Index equals, for each

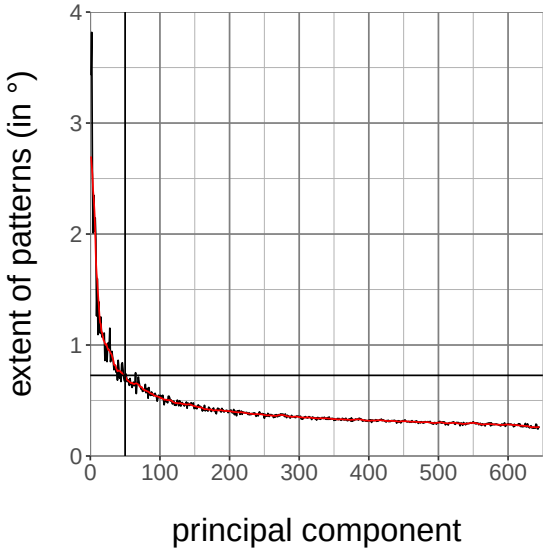


Fig. 3.3 *Extent of Pattern* for each PC for the southerly moisture transport at 700 hPa (black line). For the red line, a Gaussian smoothing is performed. The first 50 PCs are selected for this field, which correspond to an *Extent of Pattern* of 0.73° for the last PC.

predictor $\mathbf{X}_i$, the product of the coefficients $\widehat{\beta}_i$ and the correlation between the predictor $\mathbf{X}_i$ and the predictand $\mathbf{Y}$:

$$PI_i = \frac{\widehat{\beta}_i \times \sigma(\mathbf{X}_i)}{\sigma(\mathbf{Y})} \times cor(\mathbf{X}_i, \mathbf{Y}) \tag{3.8}$$

The sum of that index over all predictors equals the R^2 of the regression with all the predictors: the Pratt's Index of a predictor can be expressed as a percentage of R^2. A higher percentage means a higher contribution of a predictors to R^2. To quantify the contribution of a 2D field, we simply take the sum of Pratt's Index for each predictor composing that field, which following the notation of equation 3.7, can be simplified to:

$$PI_k = \frac{\sigma(\widehat{\mathbf{Y}}_k)}{\sigma(\mathbf{Y})} \times cor(\widehat{\mathbf{Y}}_k, \mathbf{Y}) \tag{3.9}$$

The result for each predictor is given in Table 3.2 (last row): a higher value characterises a more important predictor.

Grömping et al. (2006) and Nathans et al. (2012) have discussed advantages and limitations of Pratt's Index and other measures of relative importance. The main issue related to Pratt's Index is that negative values are possible as can be seen for some predictors in Table 3.2. These negative values don't mean that removing these variable from the regression would increase the R^2, as removing variable can only decrease it. Instead, negative Pratt's index reveals predictors that are referred to as a suppressor in regression studies Nathans et al. (2012); Smith et al. (1992). Those predictors act by suppressing the variance of other predictors. A suppressor with a high variance has a greater suppression power. Pratt's Index is generally not suitable to investigate those predictors, because correlation between suppressors and predictand is not relevant and others indices need to be looked at. However, in our context, these negative values of the Pratt's Index remain very close to zero, which suggests a low variance of the predictors, and implies that Pratt's Index offers a clear differentiation of the relative importance of each predictor, compared to other measures.

3.3.4 Time lag and causality analysis

We have so far predicted precipitation only by considering the state of the moisture transport at the start of the 3-hour precipitation accumulation period. It is possible that precipitation is also explained by an earlier state of the moisture transport due to a potential lag phenomenon. The result of the regression can also include a response of the moisture transport field to precipitation. It is therefore important to disentangle the causes and effects of precipitation. We use a distributed lag model (Saporta, 2006), where the regression is applied on the

Table 3.2 Selection of the most important predictors among the different directions of moisture transport at three altitudes: 850 hPa, 700 hPa and 500 hPa. The number of PCs for the southerly moisture transport at 700 hPa is fixed at 50; for the other fields, that number is determined so that the *Extent of Pattern* of the last PC is greater than that of the 50th PC of the southerly transport at 700 hPa. Cumulative variance explained by the selection is also given. Pratt's Index is computed on the full set of observations and given as a fraction of R^2 of the regression that considers all fields, which reaches 0.904 after validation.

Direction	Southerly			Westerly			Northerly			Easterly		
Altitude	850	700	500	850	700	500	850	700	500	850	700	500
Number of PCs	39	50	71	35	47	62	37	51	65	33	40	61
Cumulative variance explained	83%	85%	93%	83%	88%	94%	80%	85%	89%	84%	83%	89%
Pratt's Index	36%	51%	8%	1%	8%	3%	-1%	2%	1%	-3%	-5%	0%

predictors with different time lags. Due to the high correlation between the different time lags of the same predictor, multicollinearity issues arise, and the number of lags tested in the same regression is kept small. Specifically, for moisture transport, we select the start and the end of the 3-hour precipitation accumulation period, as well as three hours before the start, and three hours after the end of this period.

It is important to note that a negative lag (the predictor occurs before the predictand) does not imply a direct causality link. For example, the predictor may well be a faster response to the real direct cause of the predictand, or an indirect cause through another variable. Similarly a positive lag (the predictor occurs after the predictand) does not imply that the predictor is caused by the predictand (cf. Granger causality principles developed in Granger, 1969). Consequently, we only interpret the predictors as a potential cause (consequence) when the lag is negative (positive), as the results may differ depending on the predictors selected for the regression. A causal link can only be established after identifying and testing an underlying physical mechanism.

This causality issue can also be seen in our study. Cross-barrier flow is not a direct cause of precipitation as it relates to it through moisture convergence and condensation in the lifted air mass. We have also mentioned in the introduction that the uplift (and therefore precipitation) can be caused by convective instability. Convection results in lower troposphere convergence and thus moisture transport. If convection occurs over a large enough scale, its impact on the southerly moisture transport can be detected by our method. In our study, however, we do not distinguish whether the cross-barrier flow is a cause or a consequence of the uplift, as the two are intricately linked. Therefore, we do not investigate the specific role of instability in the generation of precipitation. Table 3.3 presents the relative importance according to Pratt's Index for the regression performed with the southerly moisture transport at 850 and 700 hPa and the four time lags selected. Pratt's Index clearly shows that the two positive time lags are much less important than the two negative, suggesting that moisture transport is predominantly a potential cause of precipitation.

3.4 Results

3.4.1 General

Table 3.2 shows that the southerly moisture transport at 700 hPa (Pratt's Index of 51%) and 850 hPa (36%) are by far the most important altitudes and directions for the prediction of the 3-hourly precipitation. In terms of time lag, the most important point in time is the start of the accumulation period (total Pratt's Index of 55%, Table 3.3), followed by three hours

Table 3.3 Selection of the most important time lags. The regression predicts the precipitation using the southerly moisture transport at 700 hPa and 850 hPa and four time lags: the start of the three-hour accumulation period of the precipitation, three hours earlier, the end of the accumulation period, and three hours later. The validated R^2 of the regression considering all the time lags reaches 0.860. The relative importance is given in the table using Pratt's Index, as a fraction of R^2.

Time steps	-3h	Start	End	+3h
Southerly moisture transport at 700 hPa	20%	34%	9%	0%
Southerly moisture transport at 850 hPa	11%	21%	0%	5%

earlier (31%). For simplification, only the start of the accumulation period is considered in the rest of the study. Figure 3.4-A presents the coefficients $\widehat{\beta}_i$ (from Equation 3.1) of the regression based only on southerly moisture transport at 700 hPa and 850 hPa at the start of the accumulation period. It shows that the highest coefficients are located along or close to the Himalayan foothills, highlighting the areas where cross-barrier moisture transport is the most important for generating precipitation. The validated R^2 of that regression is 0.832, indicating that the two predictors (Southerly moisture transport at 700 hPa and 850 hPa) explain most of the variability of precipitation. Only considering the wind reduces R^2 to 0.712, showing that although the wind explains most of the variability, moisture also plays an important role in predicting precipitation. The intercept, introduced in the statistical model ($\widehat{\beta}_0$ in equation 3.7) equals $-0.70\,\text{mm}\cdot\text{day}^{-1}$, which has the effect of offsetting the contribution of the moisture transport, suggesting that the southerly moisture transport needs to be intense enough for precipitation to occur. Note that we applied the same method to MERRA2 data (Gelaro et al., 2017), with very similar results. Overall, these results support the hypothesis that precipitation is mostly due to the condensation of moisture due to the forced uplift of a cross-barrier flow. This process is now further investigated.

3.4.2 Seasonality

Precipitation in the UIB is characterised by a bi-modal seasonality, which is well represented by ERA5 (as shown by the black line in Figure 3.5-A; Baudouin et al., 2020b). One wet season occurs in winter-early spring and is driven by extra-tropical disturbances; the other, narrower but more intense, occurs in July-August, in relation to the South Asian Summer Monsoon. Due to the difference in larger scale drivers, it is possible that the cross-barrier moisture transport is differently related to precipitation depending on the season. We

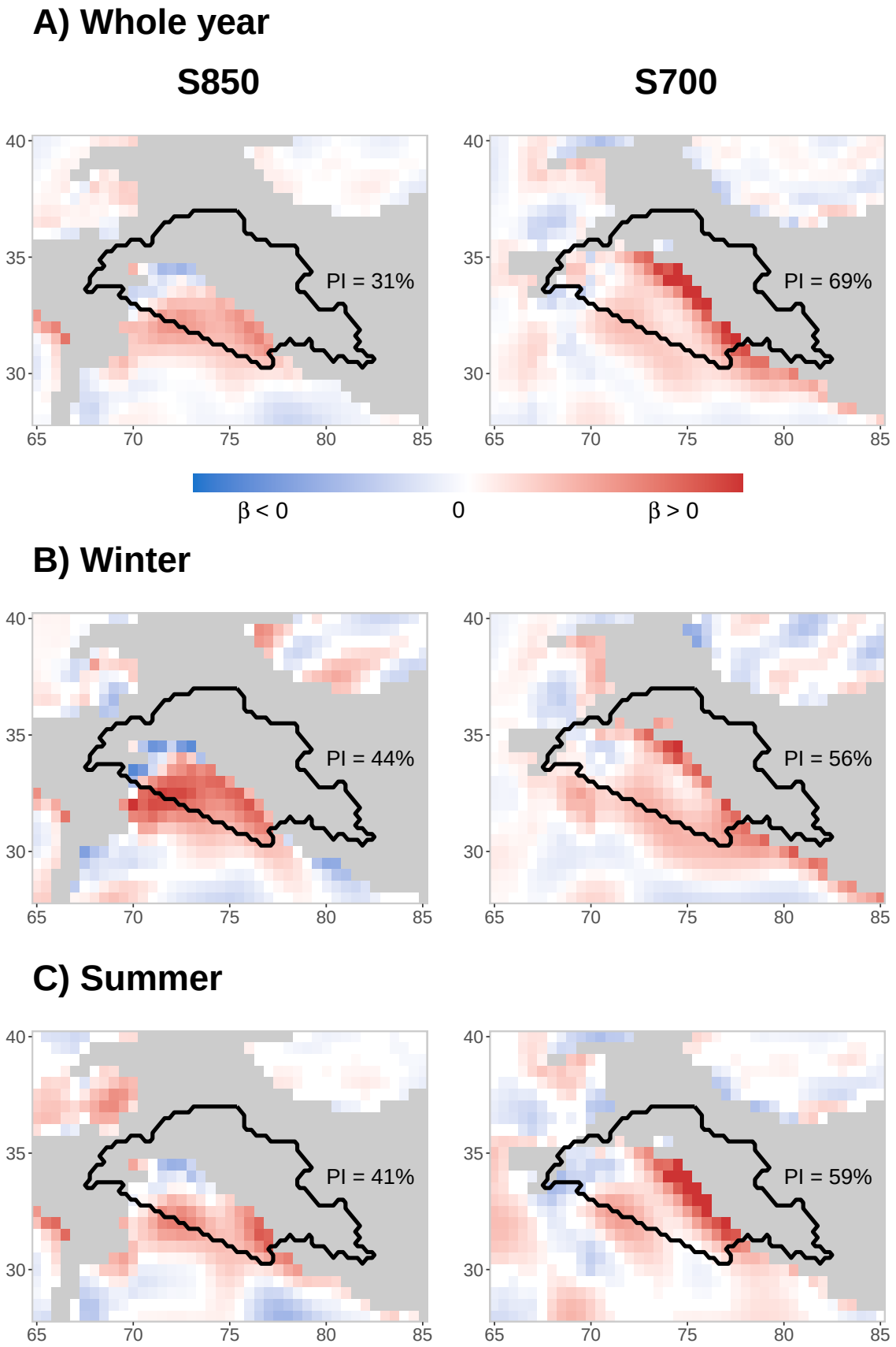

A) Whole year
S850
S700
PI = 31%
PI = 69%
β < 0
0
β > 0
B) Winter
PI = 44%
PI = 56%
C) Summer
PI = 41%
PI = 59%
40
35
30
65
70
75
80
85

Fig. 3.4 (Previous page) Map of the coefficients of the PC regression of the precipitation with the southerly moisture transport at 850 hPa and 700 hPa (S850 and S700 respectively) at the start of the precipitation accumulation period. The relative importance is given using Pratt's Index. For A, the regression considers all 3-hourly time steps (whole year) while for B and C the 3-hourly time steps are restricted to winter (December to March) and summer (July and August) respectively.

---

reproduce the same regression as in Figure 3.4-A, with the same temporal resolution, but for winter (Figure 3.4-B) and summer (Figure 3.4-C). The summer season is defined by the two wettest summer months, July and August. For winter, we consider the two wettest months February and March, but also December and January, which are generally considered in the literature on Western Disturbances (Dimri et al., 2015; Hunt et al., 2018a). A different definition of the seasons does not affect the main results.

In either season, the patterns are remarkably similar. The main changes are a re-balancing of the relative importance of the two fields, and a better coefficient of determination in winter ($R^2 = 0.922$) than in summer ($R^2 = 0.740$). We checked whether the change of relative importance of some of the other directions (westerly, easterly, northerly) and altitudes (500 hPa) for summer could explain the difference in R^2, however, there was no significant difference compared to the results presented in Table 3.2. Since no major differences are observed between the two seasonal regressions, the regression for the whole year is suited to studying the seasonality.

Figure 3.5 presents the seasonality of different variables: the prediction (3.5-A), the coefficient of determination (3.5-B), and the different contributions to precipitation (3.5-C to E). The prediction is able to reproduce almost perfectly the seasonality of precipitation (Figure 3.5-A). This quality contrasts with the seasonality of the coefficient of determination (R^2, Figure 3.5-B): it reaches a minimum in June and remains lower than it is in winter for the rest of summer. This minimum is possibly the result of increased small-scale diurnal convection during summer that produces precipitation without the need of horizontal moisture transport (e.g. influence of diurnal heating only; Houze, 2012). Yet, this process does not impact seasonal biases and the minimum value of R^2 remains high (>0.6).

Despite their similarity with the season-limited regressions, the moisture transport at 850 and 700 hPa predicting precipitation behave differently along the seasons as shown in Figure 3.5-C. Both contributions of the moisture transport exhibit two peaks that match the precipitation peaks, but their magnitudes differ. Winter precipitation is dominated by an increase in moisture transport at 700 hPa, while during summer both altitudes contribute equally to precipitation.

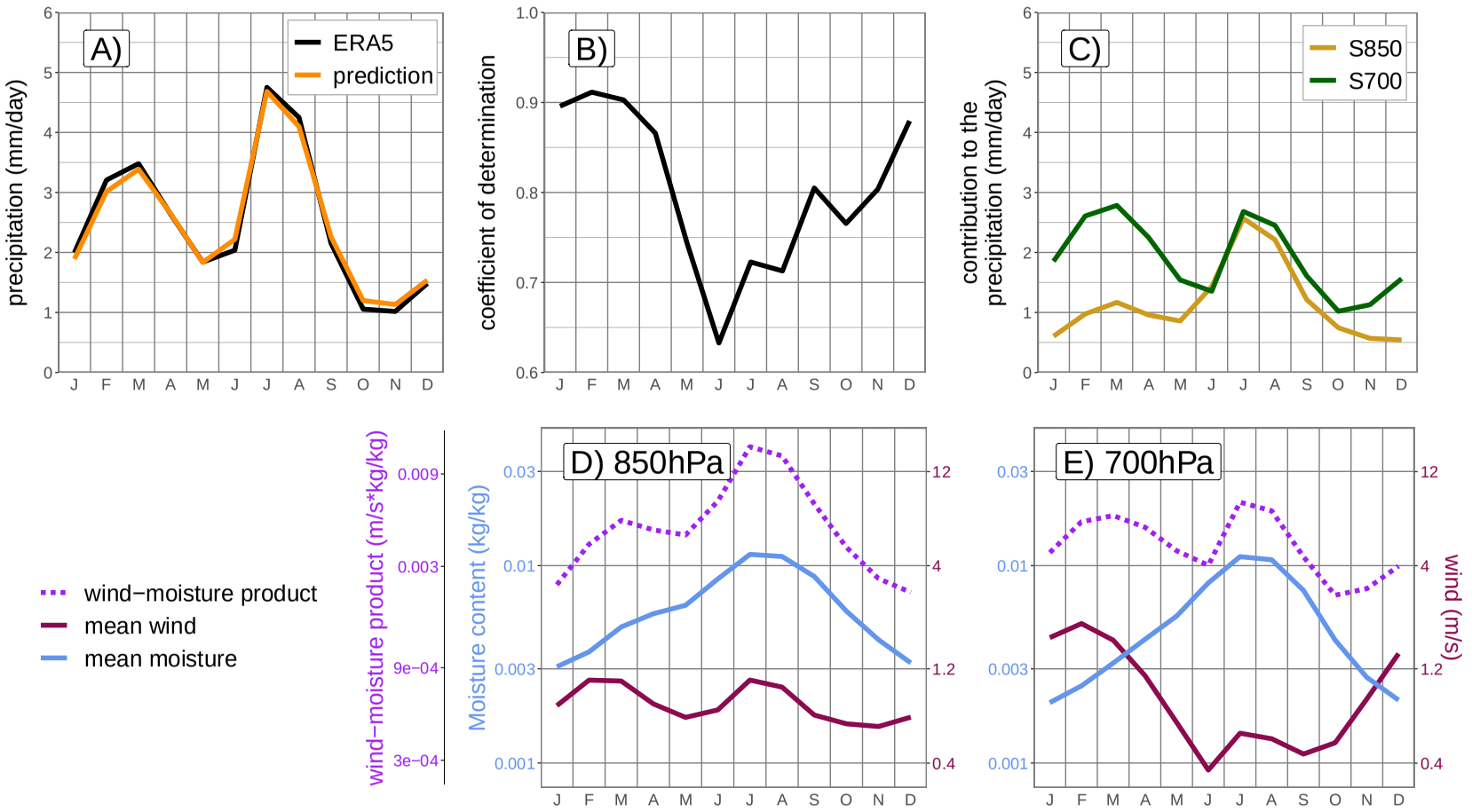


Fig. 3.5 Seasonality of different variables obtained with the same regression used in Figure 3.4-A, using ERA5 data. The monthly mean of the regression's prediction is shown in A, and compared to ERA5's modelled precipitation. The seasonality of the coefficient of determination (R^2, in B) is obtained by computing the variance of both the predictand and the residual on a specific month. The contribution of the two altitudes (S700 and S850, in C) are derived from their respective partial predictions. These contributions are further decomposed in terms of monthly mean wind and moisture (D and E).

We take the analysis further by investigating whether the seasonality in precipitation is explained by a distinct change in seasonality of the southerly wind or specific humidity. We multiply the regression coefficients (as in Figure 3.4-A) with the time series of southerly wind on the one hand and specific humidity on the other, and sum the results for each time step. The two time series obtained are representative of the distinct influence of wind and moisture, respectively, on precipitation. In Figure 3.5-D and E, we represent the seasonal cycle of those two time series on a log scale. Their product (or sum on the log scale) nearly equals the moisture transport seasonality. The small residual is due to high winds and specific humidity occurring at the same locations or time steps (e.g. spatial and temporal correlation between wind and specific humidity).

This synchronous occurrence, at the sub-daily scale, can be quite important in winter, where it explains up to 25% of the precipitation increase (i.e. the southerly winds at 700 hPa brings higher specific humidity). Yet, Figures 3.5-D and E suggest that winter precipitation is mostly driven by an increase in southerly winds at 700 hPa, while moisture only plays a role in explaining the lag between the wind peak in February and the precipitation peak in March. In contrast, the summer peak is mostly explained by an increase in moisture at both altitudes, although the wind at 850 hPa also peaks in this season, similar to winter. Both dry seasons are related to weak occurrence of southerly winds.

3.4.3 Cross-barrier wind direction

The fact that only the southerly flow relates to precipitation is quite surprising as the Himalayan range is oriented North-West to South-East. We would have expected the westerly flow to play an important role, but it is a poor predictor (total Pratt's Index of 12%, cf. Table 3.2). This result also seemingly contradicts numerous studies supporting a westward or south-westward origin of moisture, especially in the context of the Western Disturbances (Dimri et al., 2015), which are particularly evident in tracking analyses (Boschi and Lucarini, 2019; Hunt et al., 2018b; Hussain et al., 2015a; Jeelani et al., 2018). However, our study does not investigate the origin of moisture, and the PC regression result is not indicative of the mean flow in the UIB. The mean flow is given by the composite analysis we present next.

Figure 3.6 shows the composite maps of the mean wind and moisture fields when significant precipitation occurs in the following three hours ($>$ 2mm) in comparison with the absence of precipitation ($<$ 0.1mm). It shows a general southerly orientation of the wind within the study area when precipitation falls (Figure 3.6-A, C, E and G). As evident in winter at 850 hPa (Figure 3.6-C) and in summer at both altitudes (Figure 3.6-E and G), a small easterly component is also present along the foothill of the Himalayas in the study area when precipitation occurs. By contrast, a westerly component dominates in the absence of

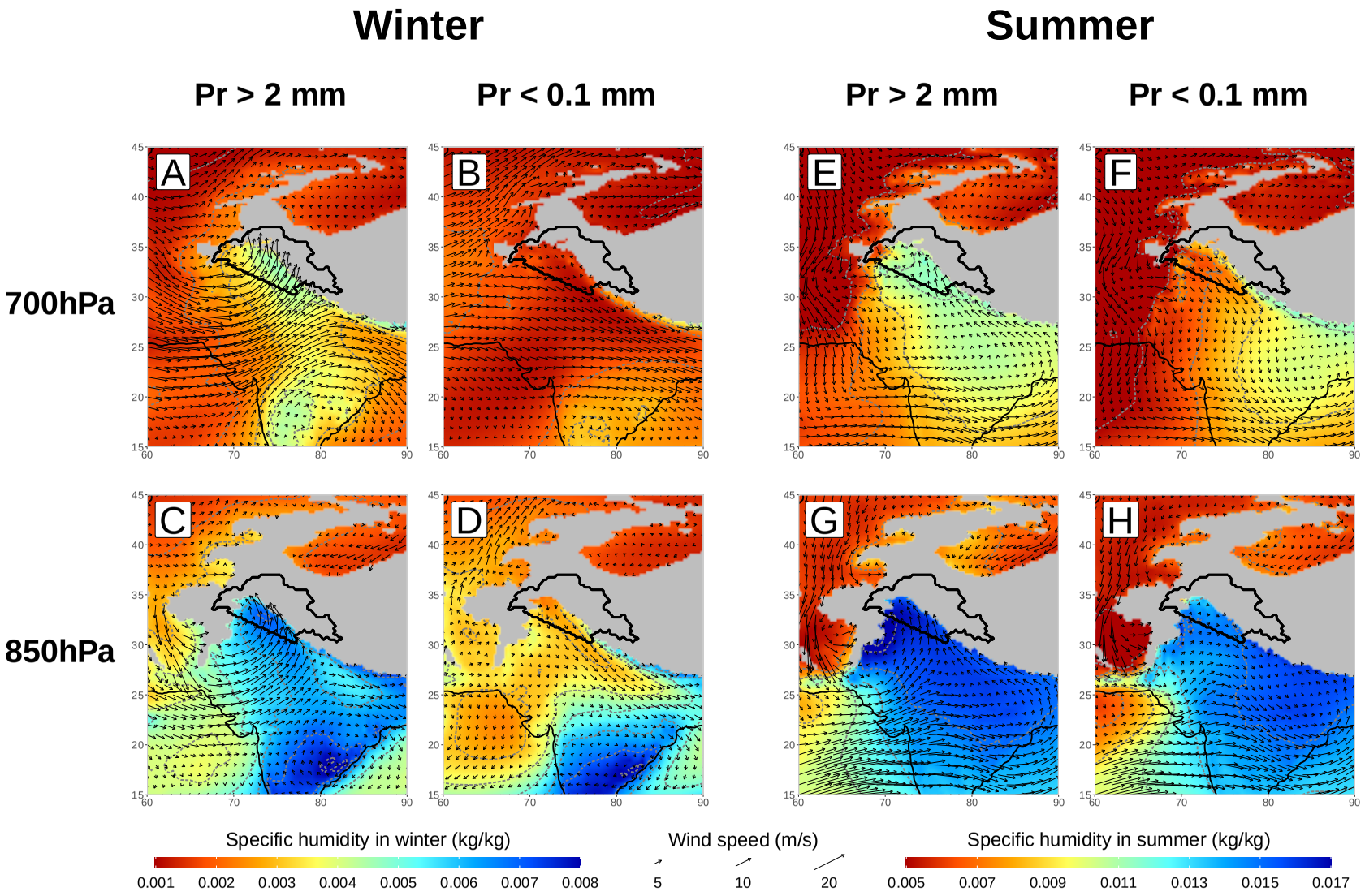


Fig. 3.6 Composite maps of average specific humidity (colour) and wind (arrows), using ERA5 (1979-2018). The maps are split by seasons (winter: December to March, in A to D; and summer: July and August, in E to H), by pressure level (700 hPa in A, B, E, and F; and 850 hPa in C, D, G, and H), and precipitation in the following 3h (above 2mm in A, C, E, and G; and below 0.1mm in B, D, F, and H). In winter (summer), there are 1029 (644) observations of precipitation above 2mm and 21220 (3368) of precipitation below 0.1mm.

precipitation (Figure 3.6-D, F and G). Several reasons explain the absence of precipitation in relation to a westerly wind, and oppositely, an easterly component when precipitation occurs.

The principal effect of topography on a flow is the deflection of that flow so that it can get around the obstacle. A horizontal deflection is particularly obvious in winter (Figure 3.6-A and B). Strong south-westerly winds occur over Rajasthan, perpendicular to the Himalayan range. Arriving close to the mountain it splits in half with a tipping point just south-east of the study area: a southerly to south-easterly branch reaches the study area while the other turns east towards the Ganges plain. A similar but weaker south-westerly winds occur over Rajasthan in summer at 850 hPa (Figure 3.6-G and H). During this season, the southerly branch heading towards the upper Indus only takes place when precipitation occurs, while the westerly branch dominates in the other case. The deflection caused by the topography should be further enhanced by the static stability of the air mass as the flow would be blocked in and below the stable layer instead of rising above the mountain range. This process, however, is not investigated further here as the static stability is not among the available variables for ERA5.

The shape of the topography also impacts the flow. There is a rapid change in the orientation of the slope of the mountain ranges in the western part of the study area, where the Hindu Kush and the Sulaiman ranges connect with the Karakoram and the Himalayas (See Figure 3.1). It forms a notch, known to be related to terrain-locked disturbances in winter (Dimri et al., 2015; Lang and Barros, 2004). When precipitation occurs, a south-easterly flow occurs parallel to the Himalayan foothills at 700 hPa or lower and is eventually trapped in that notch (Figure 3.6-C and G), and forced to rise. By contrast, if no precipitation occurs, an opposite north-westerly flow follows the Himalayas and leaves the UIB without a forced uplift (Figure 3.6-B, D, F and H).

In addition, the synoptic dynamic drives the wind flow through pressure gradients and enhances the effect of the topography. Southerly winds at 850 hPa are often triggered by a low located just to the west or south-west of the study area, as suggested by the cyclonic circulation shown in Figure 3.6-C and G. The friction, which is stronger towards the foothills, diverts the wind towards the lower pressure by geostrophic adjustment, enhancing the orographic deflection, and the convergence in the notch. Divergence is also often found at the tropopause in relation to the low at 850 hPa, particularly in winter (Hunt et al., 2018a), which sustains vertical velocities. By contrast, the topography to the north (The Pamir range and the Tibetan Plateau) prevents the formation of a low at 850 hPa there, that could enhance a westerly flow. Rather, this flow is related to the intensity of the subtropical ridge, while the topography forces an anticyclonic deflection (Figure 3.6-B, D, F and H). Overall, this synoptic context lacks the dynamics that would force the uplift of the westerly flow.

Lastly, the moisture content of the westerly and southerly flows is quite different. The southerly flow brings warm and humid tropical air. In contrast, the westerly to north-westerly flow descends after passing over the Sulaiman and Hindu Kush ranges leading to the intrusion of higher altitude continental dry air, as seen in Figure 3.6-B, D, and F (Foehn effect).

In summary, our study suggests that only the southerly component of the cross-barrier flow is important for triggering precipitation, while the mean flow varies from south-westerly to south-easterly depending on the location and season. These results corroborate the southward origin of the moisture suggested by Hunt et al. (2018b).

3.4.4 Moisture transport altitude

According to table 3.2, the moisture transport at 700 hPa is the most important predictor for explaining the precipitation variability (total Pratt's Index of 56%), followed by the one at 850 hPa (33%), while the transport at 500 hPa is much less significant (12%). We further investigate the dependency between relative importance and altitude by performing the PC

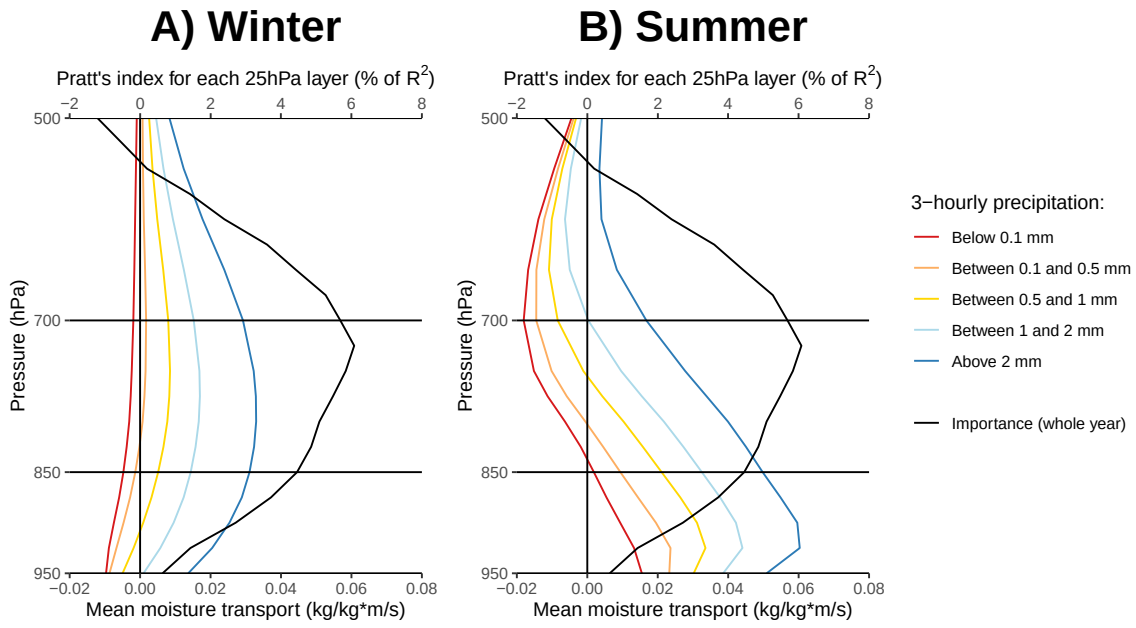


Fig. 3.7 Longitudinally averaged meridional moisture transport along 30°N and between 70° and 80°E (coloured line, bottom axis). The results are presented as a function of altitude and for different intensities of precipitation in the following 3 hours in the UIB. The results are split between winter (December-March, A) and summer (July-August, B). The black lines in A and B are identical and represent the relative importance (Pratt's Index, upper axis) of each altitude's southerly moisture transport (again, along 30°N and between 70° and 80°E) in predicting precipitation.

regression of precipitation using a vertical layer of moisture transport along 30°N (see Figure 3.1). We select as predictor all grid points above the ground between 70° and 80°E (every 0.5°) and between 950 and 500 hPa (every 25 hPa). This selection insures that we capture most of the moisture eventually reaching the UIB. The regression is performed for the whole year using the first 20 PCs. R^2 reaches 0.62, which is lower than when using horizontal layers, because the predictors considered here are further away from the foothills and the main moisture convergence area. However, R^2 remains high enough to investigate the relative importance of the different altitudes of moisture transport to produce precipitation. Figure 3.7 (black line in either A or B) shows the relative importance of each altitude using Pratt's Index. As expected, the moisture transport at 700 hPa is more important than at 850 hPa, with a peak at 725 hPa. The relative importance quickly decreases above 700 hPa and below 850 hPa, with Pratt's indices reaching values close to or below 0 at 500 hPa and 950 hPa. This behaviour justifies the selection of moisture transport at 850 hPa and 700 hPa to predict precipitation.

The relative importance of the altitude is compared to the mean meridional moisture transport across the same cross-section, which is computed for different amounts of precipitation. This composite analysis is split between the two wet seasons (Figure 3.7-A and B). The vertical structure of the mean moisture transport clearly differs from its relative importance. In winter, the moisture transport is equally strong at 700 hPa and at 850 hPa when intense precipitation occurs ($>$ 2mm per 3h over the UIB). Other altitudes above and below also see significant southerly moisture transport. During summer, the strongest moisture transport occurs at 925 hPa and quickly decreases above that altitude.

This difference in vertical structure suggests that the relative importance is not merely proportional to the mean moisture transport. The link between moisture transport and precipitation is indeed indirect. This link is first explained by moisture flux convergence along the foothills. However, not all of the moisture transported by the cross-barrier flow eventually converges. At high altitude (mostly above 600hPa), the cross-barrier flow is able to pass over the mountains with little vertical movement, minimising chances of convergence. Below, the mountain ranges effectively disturb the flow, which can converge, but can also be deflected horizontally. This deflection depends on the characteristic of the flow (static stability and wind speed) and the size of the mountain. At 850 hPa, the flow is generally weaker, and the distance to climb over the mountains larger, compared to 700 hPa. Therefore the deflection at 850 hPa is in general more important, as can be seen in Figure 3.6-A/C and E/G. Consequently, at 850 hPa, a smaller fraction of the moisture transport eventually converges and is converted into precipitation, which explains the smaller relative importance despite stronger moisture transport in Figure 3.7. Second, the link between moisture transport

and precipitation can also be modulated by the presence of moisture divergence at higher altitudes. Figure 3.4-B shows that, when no precipitation occurs, a mean southerly moisture transport is present below 850 hPa, however, above it, there is a mean northerly moisture transport: moisture escapes the UIB. In this case, higher altitudes are important not only because they provide more moisture but also because they prevent moisture from leaving the domain.

In conclusion, this analysis of the altitude's relative importance shows the advantages of the regression over the composite analysis for identifying the key components of the atmospheric circulation that explain 3-hourly precipitation variability.

3.4.5 Representation in climate models

Lastly, the method is applied to climate simulations to check the ability of climate models to represent the effect of cross-barrier winds in the UIB. We specifically use the output of three climate models from CMIP6: MRI, IPSL, and GISS, as outlined in Section 3.2.2. These datasets are only available on a 6-hourly time step, but the use of this resolution had little impact on the results discussed for ERA5. The *Extent of Pattern* used as a threshold to select the PCs for the reanalysis is too small for the resolution of the climate models. Fixing the threshold at the mean grid resolution of the datasets offers reasonable results regarding the multicollinearity issues (e.g. 1.125° for MRI, 1.875° for IPSL, and 2° for GISS).

The regression has been performed for each simulation and the coefficients are presented in Figure 3.8. The link between moisture transport and precipitation is reproduced in each simulation, as indicated by the coefficient of determination (R^2) of the regressions, which is in the range of 0.8 for the three simulations. Despite various spatial resolutions, they all represent the highest coefficients along the Himalayas. Some discrepancies with ERA5 are notable in the relative importance of each predictor. Moisture transport at 700 hPa is more important for IPSL and MRI, while in GISS, precipitation is more related to moisture transport at 850 hPa. This behaviour may be related to the representation of the relief and the model's latitudinal resolution (1.125° for MRI, 1.25° for IPSL, and 2° for GISS).

The seasonality of precipitation and its different contributions is represented in Figure 3.9, while the seasonality of R^2 is shown in Figure 3.10. The seasonal cycle of precipitation is similar in the three models, but is very different from reality, as was noted from previous versions of the models used for CMIP5 (Palazzi et al., 2015). The wet season in summer is absent, instead the period June to September is the driest. Winter precipitation, on the other hand, is more intense and lasts longer from October, or November to April, or May, depending on the model, with a peak in April, instead of March. The overall annual precipitation is 5% (for IPSL) to 15% higher (for MRI) compared to ERA5.

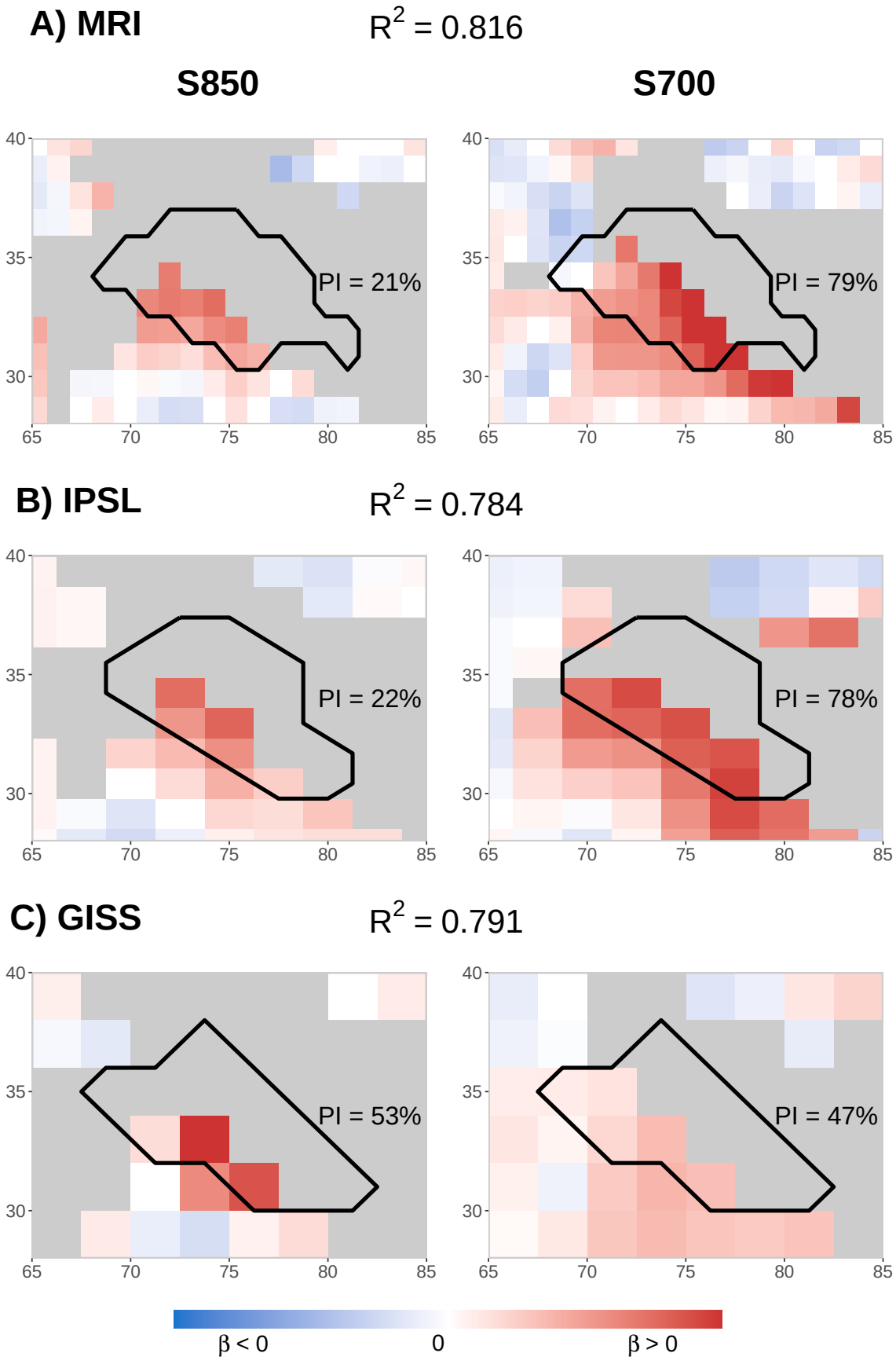

A) MRI
R^2 = 0.816
S850
S700
PI = 21%
PI = 79%
B) IPSL
R^2 = 0.784
PI = 22%
PI = 78%
C) GISS
R^2 = 0.791
PI = 53%
PI = 47%
40
35
30
65
70
75
80
85
β < 0
0
β > 0

Fig. 3.8 (Previous page) Same as Figure 3.4-A, but with data from the climate models: MRI (A), IPSL (B) and GISS (C). The study area used to compute the average of precipitation is adapted to the resolution of each dataset (black contour). The R^2 of the regression with moisture transport at both 850 and 700 hPa is also given.

The regression of precipitation helps to understand some of the biases in modelled precipitation. First, the underestimation of summer precipitation coincides with a drop of the predictive skill: the R^2 of the regression is around 0.3 or below in July (Figure 3.10). This drop suggests that large-scale cross-barrier winds are no longer the main trigger of summer precipitation in the simulations. The drop may also be related to the coarser resolution which increases the importance of sub-grid precipitation. Second, the moisture content related to precipitation is underestimated for all seasons and all models compared to ERA5 (Figure 3.9-G to L). In winter, this underestimation is more than compensated for by stronger southerly winds which peak at both altitudes over an extended winter season. In summer, by contrast, the wind does not peak at 850 hPa in the simulations as it does in ERA5 and reaches a minimum instead.

In summary, climate models produce precipitation in the UIB for the right reason: the forced uplift of a cross-barrier flow and the condensation of the moisture within. Yet, the seasonality of this moisture transport is largely incorrect leading to the precipitation biases. We show that the synoptic-scale circulation is responsible for these biases, rather than the moisture processes.

3.5 Conclusion

A method based on statistical regressions and principal component analysis (PC regression) is used to investigate the link between a given variable (predictand) and 2D fields that characterise the atmospheric state (predictors). More specifically, the method evaluates the strength of this link and exposes the pattern of the 2D fields that explain the predictand. Those patterns are useful for interpreting the link in terms of physical processes, while lag regressions can be used to disentangle cause and effect of the predictand's variability. This study also stresses the advantages over a composite analysis, which is often used in atmospheric science. Overall, the PC regression allows for a comprehensive analysis of the causality links and can be applied to different contexts other than the one investigated here.

This study focuses on the causes of precipitation variability in the upper Indus River Basin. Only the mean precipitation over that area is considered, as no dataset reliably represents

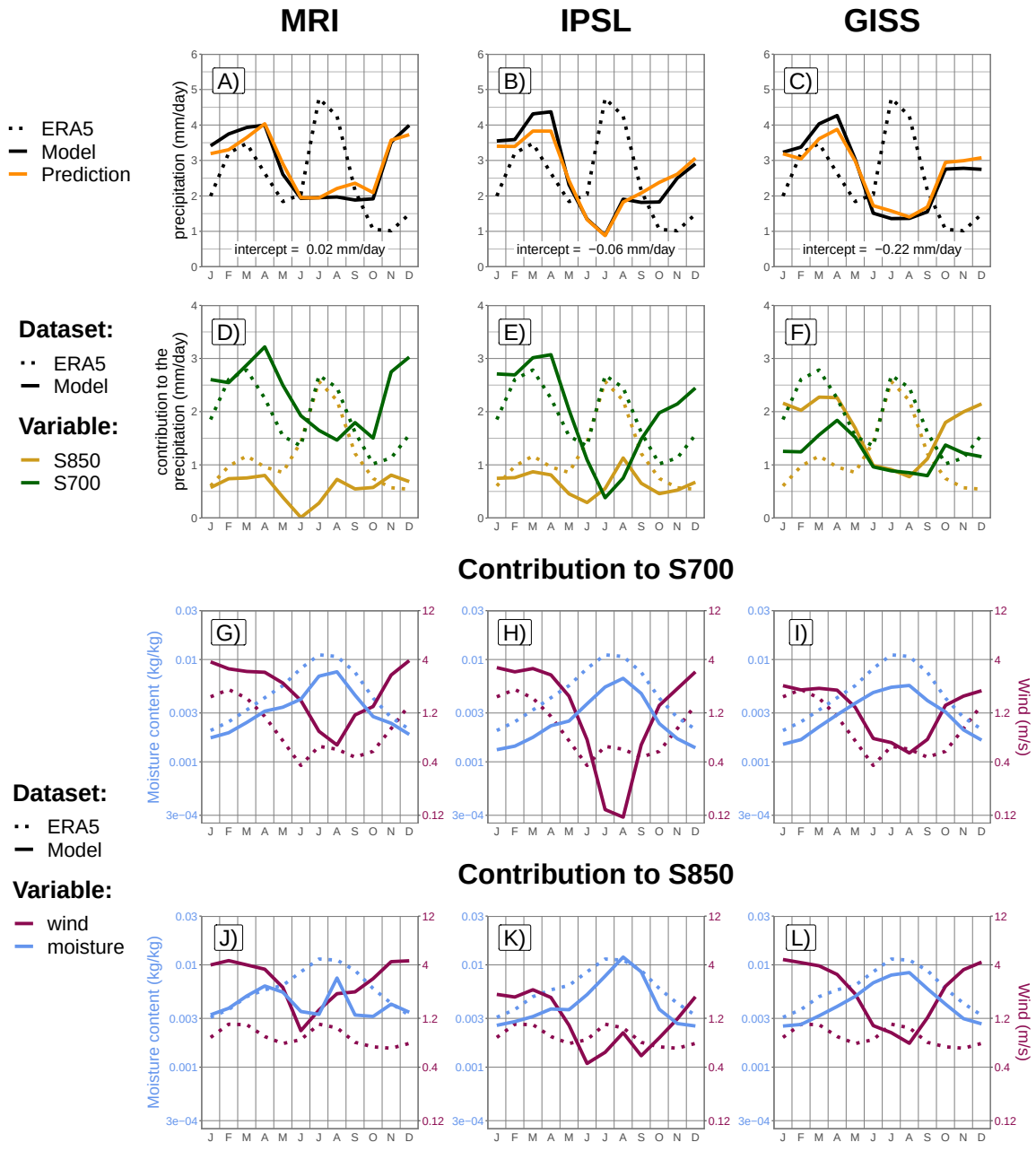


Fig. 3.9 Seasonality of the precipitation, the prediction, and its different contributions for the climate simulations (MRI: A, D, G, and J; IPSL: B, E, H, and K; GISS: C, F, I, and L). Model's data (thick line) are compared to ERA5 data (dashed line) as showed in figure 3.5).

the fine scale patterns of precipitation (Baudouin et al., 2020b). The main result is that horizontal moisture transport successfully predicts over 80% of the precipitation variability of which southerly moisture transport at low levels (between 700 hPa and 850 hPa) along the Himalayas is the dominant contributor. This fact demonstrates that precipitation in the upper Indus basin is mainly caused by a forced uplift of a cross-barrier flow. Comparison with mean moisture transport through a composite analysis complements the PC regression by connecting its result to the larger-scale circulation and the synoptic drivers of precipitation. However, important discrepancies arise from the vertical structure between mean and relative importance of moisture transport, which suggest complex, altitude dependent, interaction with the relief. Particularly, the mountain ranges and their specific shape play an important role in the channelling, trapping and eventually uplifting of moisture coming from the south of the study area. A finer-scale analysis would be useful to further investigate this relationship. We also suggest that a complex interaction between cross-barrier flow and the stability of the air mass provides the energy that sustains vertical velocities and moisture convergence. However, the method and data used here are not suited for such an analysis.

The link between moisture transport and precipitation for both winter and summer are strikingly similar despite differences in the synoptic drivers of precipitation, proving the importance of cross-barrier transport of moisture in both seasons. Nonetheless, in summer, moisture transport explains a lower fraction of the precipitation variability, which suggests that additional factors influence precipitation, such as small-scale convection triggered by diurnal differential heating.

The prediction of precipitation is decomposed into the contributions of the different altitudes of moisture transport and, from there, into the contributions of wind and moisture.

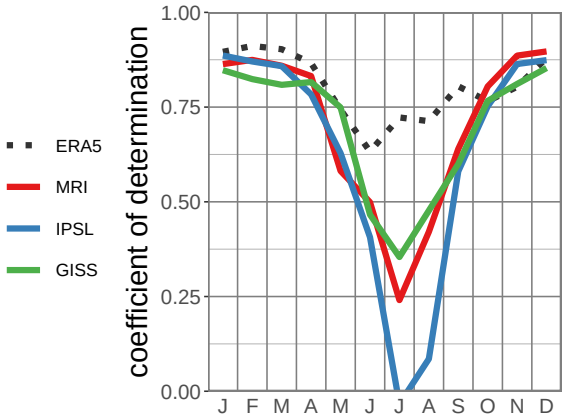


Fig. 3.10 Seasonality of the R^2 as computed in Figure 3.5-B for the climate models (thick lines) compared to ERA5 (dashed line, as in Figure 3.5-B).

These contributions show further differences between seasons. Winter precipitation variability is solely driven by moisture transport at 700 hPa, while in summer, moisture transport at 850 hPa and 700 hPa have similar contributions. In addition, the winter peak is driven by an increase of mean southerly wind while the mean moisture is at its minimum. The increase of moisture starting from January explains the delay between the peak in wind in February, and the peak in precipitation in March. By contrast, the summer wet season is mostly explained by the increase in moisture content, while the mean southerly wind also exhibits a secondary peak. Both dry seasons are explained by the absence of occurrences of southerly winds. The decomposition in altitude and wind/moisture contributions offer further opportunities to investigate the variability of precipitation at longer timescales through the variability of these contributions.

Lastly, the method provides insight into the reasons why climate models misrepresent precipitation seasonality. The CMIP6 climate models selected for this study represent similarly overly wet winters and dry summers. Despite precipitation biases and relatively coarse spatial resolution, the climate models are able to reproduce a relationship between moisture transport and precipitation similar to the reanalysis. The decomposition of precipitation into the various contributions suggests that both winter and summer biases are related to an anomalous seasonality of the mean southerly wind, while moisture is relatively well simulated. These anomalies suggest that the synoptic-scale circulation, rather than local moisture processes, is the main factor explaining precipitation biases in these climate simulations. Further analysis of the wind and moisture variability is needed to narrow down and correct climate model biases and build confidence in simulations for past and future climates.

This chapter has developed a method to evaluate the trigger of precipitation at the subdaily scale in climate models. It has also discussed the representation of the precipitation seasonality. To further understand the origin of the seasonal biases, a better understanding of the synoptic driver is needed, and will be developed in chapter 4.

Chapter 4

Synoptic variability of Western Disturbances and precipitation in the UIB

4.1 Introduction

As discussed in Chapters 1 and 2, the Upper Indus River Basin (UIB) is specific compared to the rest of the Indian subcontinent for the large amount of precipitation it receives outside of the summer monsoon season (56% or 505mm between October and May for ERA5 on the period 1979 and 2018). Much of this precipitation falls as snow at altitude during the coldest part of the season (Chapters 1 and 2, Dahri et al., 2018; Hewitt, 2011). The precipitation and the snowmelt later in the season are key to mitigating the seasonal drought that occurs over most of South Asia before the arrival of the summer monsoon (Chapter 1; Dimri et al., 2015; Rana et al., 2015; Singh et al., 2011).

Observational datasets underestimate the total precipitation amount leading to important uncertainty (Chapter 2; Baudouin et al., 2020b; Dahri et al., 2018). Yet, reanalyses, particularly ERA5, represent precipitation variability at various timescales similar to that in observations, which suggests they are able to represent precipitation drivers as well (Chapter 2) and can be used to investigate those drivers (Chatper 3; Baudouin et al., 2020a).

The drivers of winter precipitation differ from those of the summer monsoon (Chapter 3) and have been discussed since the mid-twentieth century (Malurkar, 1947; Mull and Desai, 1947, cf. Dimri et al. 2015). Winter precipitation events in the UIB are related to the passing of extra-tropical, synoptic-scale, disturbances often originating from the Mediterranean Sea, the Black Sea, or the Caspian Sea, and referred to as Western Disturbances (WDs, Dimri

et al., 2015; Hunt et al., 2018a). Typically up to six or seven events occur per month (Hunt 2018), although lower frequencies have been observed (Cannon et al., 2015; Dimri, 2013a). A WD is characterised by a maxima of vorticity travelling in the Sub-tropical Westerly Jet (SWJ, Dimri et al., 2015; Hunt et al., 2018a), the latter establishing just south of the UIB region in winter (Ahmed et al., 2019; Schiemann et al., 2009). Several studies have used tracking algorithms to describe the statistics of WDs but with limited insight on the precipitation variability (Cannon et al., 2016a; Hunt et al., 2018a; Syed et al., 2010).

As discussed in Chapter 3, the primary trigger of precipitation is the forced up-lift of a cross-barrier flow and the condensation of the moisture within it (Baudouin et al., 2020a). Numerous numerical studies have investigated WD characteristics during specific events and confirmed the importance of relief convergence (e.g. Dimri, 2004; Dimri and Niyogi, 2013; Krishnan et al., 2018; Thomas et al., 2018). Despite this abundant interest, a general understanding of the relationship between WD characteristics and precipitation is still lacking.

Chapter 3 highlighted the importance of both wind and moisture to explain precipitation. Yet, those two components exhibit a different temporal variability, which suggests different drivers. Using ERA5 reanalysis (Section 4.2), this chapter further develops the method used in Chapter 3 (Section 4.3). It then describes the characteristics of WDs in relation to precipitation (Section 4.4), the behaviour of the Subtropical Westerly Jet (Section 4.5), the origin and displacement of moisture (Section 4.6) and finally discuses the causes of precipitation seasonality (Section 4.7).

4.2 Season, data and study area

The timeseries of precipitation considered here is defined by a 3-hourly average over the UIB (cf. black contour in Figure 4.1-A). This study area is the same as the one used in Chapters 2 and 3. Both precipitation and atmospheric variables are derived from ERA5 reanalysis, at a 3-hourly intervals and 0.5° resolution, as in Chapter 3. ERA5 data provide extrapolated values on pressure levels below the model surface which should not be used in the analysis. Consequently, grid points where the minimum geopotential is above the model surface were deselected (instead of using an arbitrary pressure threshold as in the previous chapter, Section 3.2.2).

All analyses are performed over an extended winter season that spans from October to May. Similar studies on WDs generally considered a shorter winter period to avoid the dry intermediate seasons (e.g. December-February in Midhuna et al. 2020 or December-April in Hunt et al. 2018b). However, a longer season enables the analysis of precipitation drivers

during these intermediate seasons and their similarities and differences to that of the wet winter season. This selection is further discussed in Section 4.7.

4.3 Computing wind and moisture contribution to precipitation

Chapter 3 demonstrated the ability of principal component (PC) regression to analyse the precipitation variability in terms of moisture transport, defined as the product of wind with specific humidity. The same method is used again here, with small differences, to both simplify and test the robustness of the method. The sensitivity of the coefficient of determination (or explained variance) R^2 to those changes is tested and the results are explained below.

Chapter 3 made use of southerly moisture transport at 850 and 700 hPa to predict 3-hourly precipitation over the whole year. The same is done here, but over the period October to May. Using the first 50 principal components of moisture transport at 700 hPa, and the first 40 at 850 hPa, the R^2 of that regression reaches 0.899.

Furthermore, the timeseries of the predictors have been selected over a smaller area: between 18 and 36°N and between 67 and 85°E, still using a 0.5° resolution (cf. the extent and resolution in Figure 4.1-A, compared to Figure 3.2). This area was chosen in order to zoom in on the Himalayan foothills, and particularly to remove the influence of grid points located beyond the mountain range, such as those in the Tarim basin to the north-east, the Amu-Daria basin to the north-west and the Helmand basin to the south-west. The number of principal components selected for the regression is recomputed using the same threshold for the Extent of Pattern (EP) as it was defined in Chapter 3: the last selected component has an EP above 0.73°. A total of 27 components have been selected for the moisture transport at 700 hPa, and 22 at 850 hPa. R^2 is only reduced from 0.899 to 0.889, showing the weak impact of the area reduction.

One of the main results of Chapter 3 was the fact that the most important contributor to winter precipitation variability is the moisture transport at 700 hPa, in agreement with the maximum wind convergence altitude found in Hunt et al. (2018a). Thus, only this altitude is considered here to simplify the analyses. The R^2 of the regression that only considers this predictor is reduced to 0.854, which remains close enough to the quality of the regression that considers the two altitudes.

Small changes have also been made to the PC regression. The Principal Component Analysis (PCA) performed in Chapter 3 was based on standardised time series, as this is

usually done. However, this leads, once the regression is performed, to an over-representation of grid-points whose time series have low variance. Here, the PCA is instead performed on predictors that are only centred (i.e. the mean is removed). The number of selected principal components is reduced to the first 22, as the new components have smaller EP. The R^2 of the regression is slightly reduced to 0.846.

Furthermore, in Chapter 3, a linear statistical model was used. This method has the disadvantage of enabling negative values for the predicted precipitation, leading to an overestimation of the variance. This deficiency can be corrected by fixing negative values to 0 in the statistical model, and solving the regression using an iterative optimisation method (Brent, 1973). A condition on the absence of bias is also added to the model. It results in a small increase of the R^2 to 0.854.

Finally, in addition to only considering southerly moisture transport as in Chapter 3, both meridional and zonal components are simultaneously considered here, allowing a better visualisation of cross-barrier moisture transport. The PCA combines both components and the first 46 PCs are selected. The resulting R^2 is further increased to 0.878. Figure 4.1-A shows the coefficients of that regression: the arrows are made using the coefficients of each component, while the colour is proportional to the length of the arrow. The arrows highlight the moisture convergence zone along the Himalayan foothills that triggers precipitation.

Similar to Chapter 3, the prediction is split between the contribution of wind and moisture, using the weighted spatial averaging. The wind contribution (labelled hereafter $W700$) is computed by multiplying the time series of meridional and zonal wind with the respective meridional and zonal regression coefficients and summing the result. For the moisture contribution (labelled hereafter $Q700$), the time series at each location are weighted with the euclidean norm of the coefficients of both meridional and zonal moisture transport (i.e. colour scheme in Figure 4.1-A). This operation corresponds to a weighed spatial averaging. Fixing negative values of $W700$ to 0, the product of $Q700$ and $W700$ has an R^2 of 0.844 with precipitation, close to the R^2 of the regression with moisture transport (0.878), despite the spatial averaging. Note that when considering the product of $Q700$ and $W700$, the intercept of the regression, which equals $0.14\,\mathrm{mm}\cdot 3\mathrm{hours}^{-1}$, cannot be taken into account.

Figure 4.1-B shows the product of $Q700$ and $W700$, against precipitation. A non-linearity can be seen as precipitation increases quicker than the product. This non-linearity is not present when including moisture transport at 850 hPa in the PC regression (As in Chapter 3, Section 3.4.1). This behaviour can be explained by the fact that moisture transport at lower altitude only increases and affects precipitation when strong moisture transport is already present at 700 hPa (cf. Figure 3.7). An ad hoc quadratic fit of the precipitation with the product of $Q700$ and $W700$ captures the non-linearity and slightly increases the R^2 from

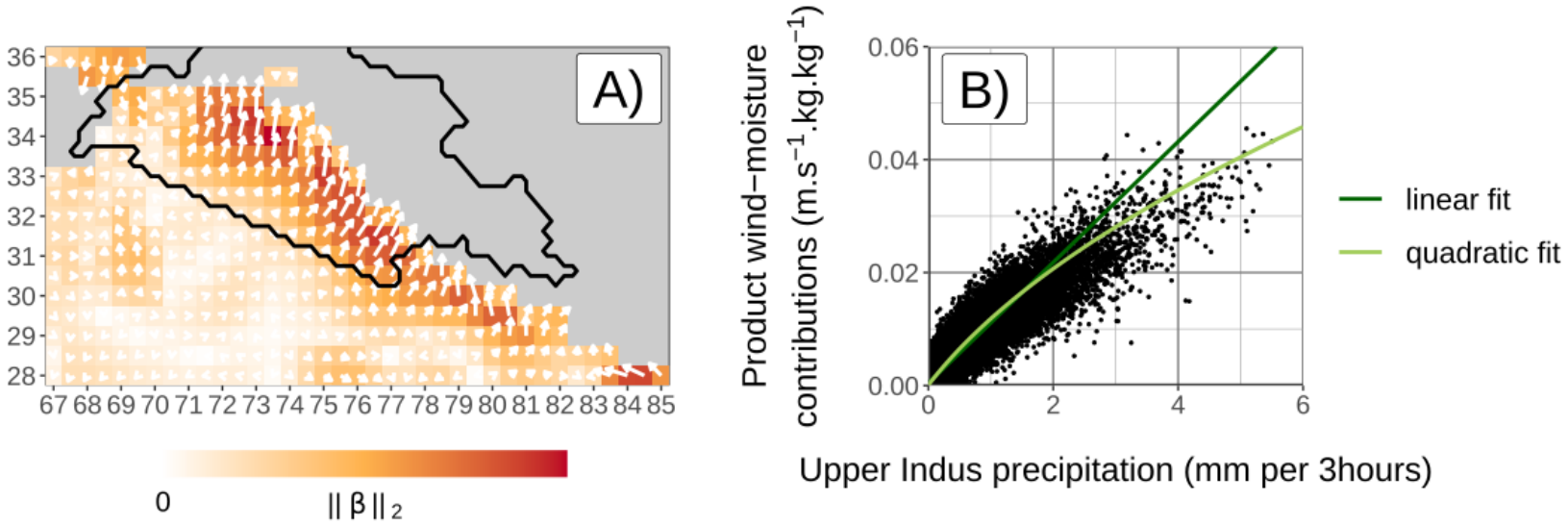


Fig. 4.1 Results of the wintertime regression of precipitation with meridional and zonal moisture transport at 700 hPa. Precipitation is defined as the 3-hourly averaged over the UIB (black contour Panel A). Panel A shows the coefficients β associated with each predictors. The arrows indicate the most efficient direction of moisture transport, while the arrows length and the colours indicate the total weight of moisture transport (i.e. the Euclidean norm of the coefficients β from both zonal and meridional components, $\sqrt{\beta^2_{zonal} + \beta^2_{meridional}}$). The exact values of the coefficients are not interpretable and are therefore not indicated. Panel B is the scatter plot of the product of wind and moisture contributions against precipitation, with the green lines representing two types of fit.

0.844 to 0.853. Hence, the product of $Q700$ and $W700$ is a good predictor of precipitation and each contribution can be investigated separately.

The analysis of the R^2 values suggests that the PC regression is robust to the small changes in the method, and that the variability in wind and moisture contribution can be analysed to study the variability in precipitation.

4.4 Relating wind contribution to Western Disturbances

4.4.1 Composites

A WD is defined as a maximum in relative vorticity (Hunt et al., 2018a), or a minimum in geopotential height (Cannon et al., 2016a; Syed et al., 2010), generally located at high altitude, close to the tropopause between 325 hPa (Hunt et al., 2018a) and 200 hPa (Midhuna et al., 2020). These anomalies drive a cyclonic circulation around the centre: the southerly advection ahead of the WDs interacts with the relief and produces precipitation (cf. Chapter 3, Figure 3.6-A and C; Dimri, 2004; Hunt et al., 2018a). Yet, the convergence triggering precipitation occurs at a much lower altitude, around 700 hPa and below (Chapter 3 Section 3.4.4; Hunt et al., 2018a), which suggests that the cyclonic circulation extends towards the surface. Alternatively, Dimri and Chevuturi (2014) propose that the WD interacts with a pre-existing low-level cyclonic circulation located over the Thar desert, similar to the heat low present in the area during summer (Bollasina and Nigam, 2011). The interaction between an upper- and a lower-level low is known as baroclinic interaction, and is a key process in the growth of extra-tropical disturbances (Malardel, 2005). This section will further investigate the relation between WDs and lower level cross-barrier flow, as well as other parameters modulating the interaction between the two.

The relation between WDs and $W700$, the wind contribution to precipitation as computed in Section 4.3, is introduced using a composite analysis. The composite is computed by selecting 10% of the 3-hourly time steps with the highest $W700$ (i.e. $W700$ above the 90$^{\text{th}}$ percentile: 2.68 $\mathrm{m\,s^{-1}}$). Those time steps are responsible for 50% of the precipitation during winter. Note that this selection will be re-used for most composite analyses throughout this chapter. The anomalies of geopotential height, or of other variables, are computed by removing the four first harmonics of the seasonal cycle. This approach ensures, for example, that the monthly mean anomalies of geopotential height at 300 hPa are below 100 $\mathrm{m^2\,s^{-2}}$.

Figure 4.2 represents the geopotential height anomaly at 300 hPa as well as the anomaly of geopotential thickness between 500 hPa and 300 hPa. The thickness anomaly is directly indicative of temperature anomalies in that layer under the hydrostatic equilibrium. During

the time steps considered (Figure 4.2-B, "No lag"), a mean minimum of geopotential anomaly is present at 36°N 66°E, just north of the Hindu Kush, and north-west of the UIB. The lead/lag analysis (Figure 4.2-A and C) reveals the eastward motion of the disturbance, along 36°N, which therefore fits the definition of the WDs.

The WD in Figure 4.2 has a cold core, which implies that the geopotential height anomaly increases with altitude (This will be further explored in the following Figure 4.3). The centre of the cold core is, however, slightly displaced to the north-west of the geopotential anomaly minimum, which indicates baroclinicity. Both of these characteristics are in agreement with the literature (Dimri et al., 2015; Hunt et al., 2018a).

The minimum of the geopotential anomaly is also accompanied by a weaker maximum located over the Tibetan Plateau, at the edge of the figure (Figure 4.2-B). This second centre of action is already present over the UIB two days before the wind peak (Figure 4.2-A). The two centres form part of a wave train characterising the development of WDs (Dimri, 2013a;

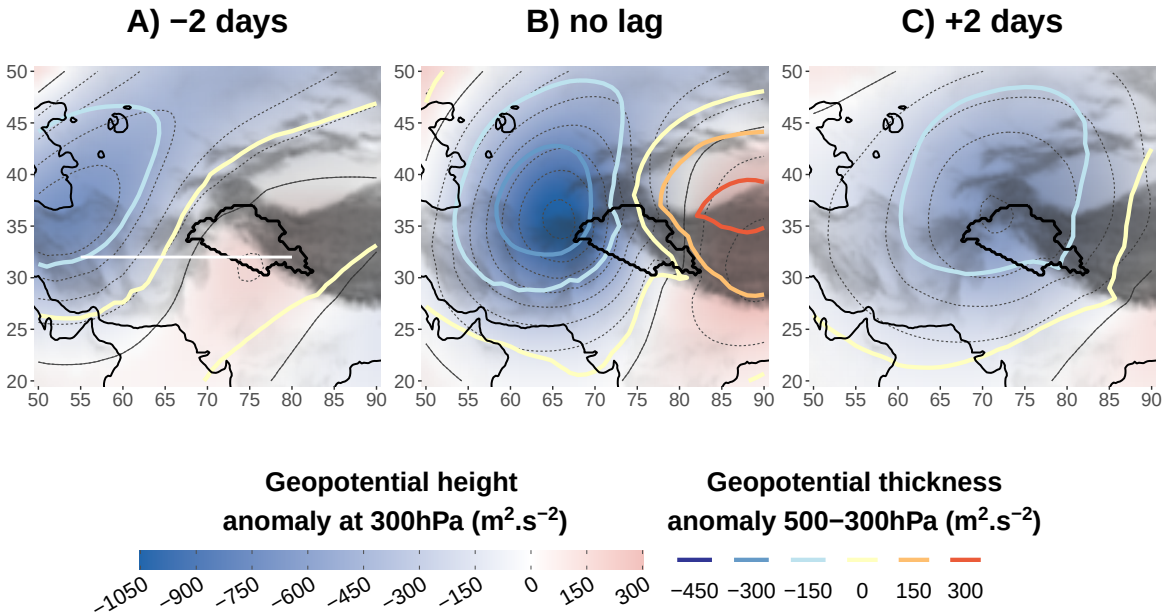


Fig. 4.2 Composite maps of geopotential height anomaly at 300 hPa (colour shading, thin contour lines every 150 $m^2 s^{-2}$) and geopotential thickness anomaly 500-300 hPa (thick contour lines). Values are an average based on the 10% highest values of the 700 hPa wind contribution during winter ($W700$, panel B). For panel A (C) a two days lead (lag) is applied to the selection. The white line in panel A indicates the cross-section in Figure 4.3. Relief (grey shading) and coastline are based on ERA5 data. The UIB is indicated in each panels. Non-significant anomalies at the level 95% are shown in white (result of a t-test on the means).

Hunt et al., 2018b). These previous studies have not stressed the importance of this second extrema, which increases the zonal gradient of geopotential to the east of the WD and thus strengthens the southerly wind, and is key to the relief interaction.

The warm core (i.e. the positive geopotential thickness anomaly) to the east of the WD is displaced compared to the geopotential anomaly maximum (Figure 4.2-B). This is likely the result of a Foehn effect on the leeward slopes of the mountain ranges: higher temperature is both related to latent heat release and adiabatic compression. Geostrophic adjustment in response to the outflow, parallel to the temperature anomaly gradient, could be another reason and will be further discussed in Section 4.5.

Figure 4.3 represents a vertical cross-section of the troposphere along 30°N. The section is located south of the centre of the WD, so it cuts across the Indus Plain between 70°and 78°E, just south of the UIB (cf. white line in Figure 4.2-A), where the southerly advection originates. Different anomaly fields are investigated: geopotential height and temperature anomalies in Figure 4.3-A to C, and meridional wind and specific humidity anomalies in Figure 4.3-D to F, although specific humidity will only be discussed in Section 4.6.

The maximum anomaly of geopotential height is located around 300 hPa (Figure 4.3-A and B), in agreement with the literature (Hunt et al., 2018a; Midhuna et al., 2020). There is no evidence of a separate cyclonic circulation at low level, however (Figure 4.3-B), nor a pre-existing circulation for the day before (Figure 4.3-A), as mentioned by Dimri and Chevuturi (2014) and Dimri et al. (2015) among others. Instead, the figure shows that the lower level circulation is simply a weak extension of the anomaly at 300 hPa. This downward extension of the cyclonic circulation is also evident in the meridional wind field (4.2-D). However, Dimri and Chevuturi (2014) might have been referring to fast-moving small-scale eddies that could circulate at low level above the Indus Plain and that are not detected by the composite analysis. Regardless, there is no evidence of a baroclinic interaction between upper-level and lower-level disturbances that characterises most intense cyclonic activity elsewhere (Dacre et al., 2012; Malardel, 2005). Instead, WDs are immature baroclinic waves (Hunt et al., 2018a) that activate through the interaction with the relief.

The vertical black lines in Figure 4.3-A and B indicate the location of the maximum anomaly for each altitude and mostly correspond to the contour of zero meridional wind anomaly (Figure 4.3-C and D respectively). The lines exhibit a westward tilt with the altitude, which is a sign of baroclinicity (Dimri and Chevuturi, 2014; Hunt et al., 2018a). This baroclinicity is also evident from the temperature anomalies: colder air is present to the west of the WD's centre rather than to the east, as was also noted in 4.2-B. It is this asymmetry between temperature and geopotential height anomalies that causes the tilt. The asymmetry further increases at lower levels, below 700 hPa, where cold continental dry air is advected

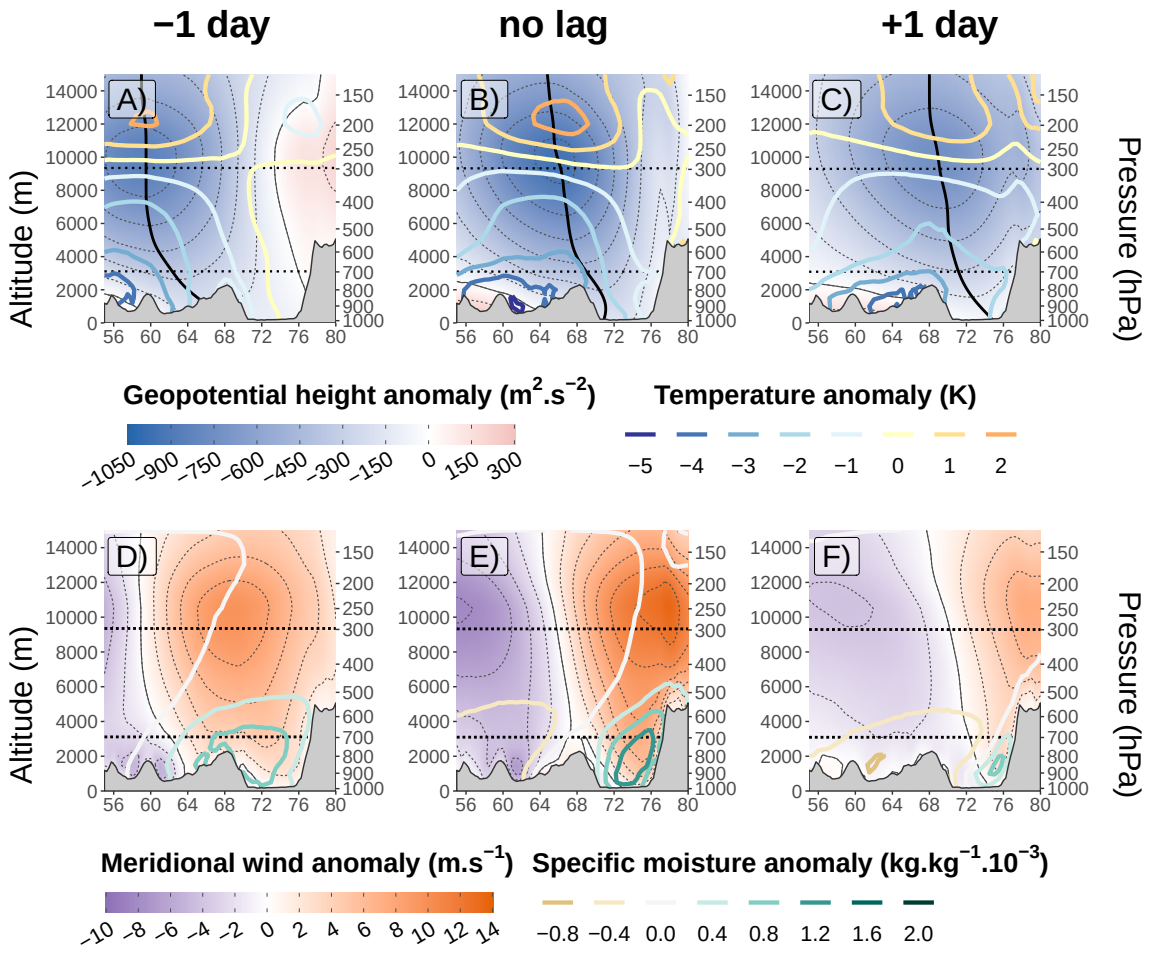


Fig. 4.3 Composite cross-section along 30°N of anomaly of geopotential height (colour shading, thin contour lines every 150 $m^2\,s^{-2}$) and temperature (thick contour line) in panels A to C, and meridional wind (colour shading, thin contour lines every 2 $m\,s^{-1}$) and specific moisture (thick contour line) in panels D to F. Values are an average based on the 10% highest values of $W700$ (B, E). For panel A and D a one day lead is applied to the selection, while a one day lag is applied for panel C and F. The dotted horizontal lines represent the level 700 and 300 hPa, the altitudes for the cross-barrier moisture transport and minimum geopotential anomaly, respectively. The vertical black line in A, B, and C represents the longitude of minimum geopotential anomaly as a function of altitude. The grey shading represent the relief as in ERA5.

from Siberia and Central Asia at the rear of the WD (Figure 4.3-A, Yadav et al., 2012). The Sistan plain in eastern Iran opens a north-south valley between the Hindu Kush and that part of the Iranian plateau around 62°E in the cross-section that funnels the cold air flow (Figure 4.3-B and D). Surprisingly, negative temperature anomalies are also present in the Indus Plain, despite southerly winds bringing in warmer air (4.2-B). This appearance may be related to the adiabatic cooling from the geostrophic adjustment to the baroclinic instability. Closer to the surface precipitation evaporation and reduced solar radiation could further explain the cooling.

The downward propagation of the cyclonic circulation explains the presence of a zonal gradient of geopotential in the Indus Plain. This gradient forces a southerly geostrophic wind at 700 hPa that converges in the UIB. The zonal gradient is already present one day before the wind peak at 700 hPa, despite the mean WD centre located at about 1500 km to the west, owing to the tilt of the circulation and the presence of positive geopotential anomalies towards the Himalayas. Hence, southerly winds, relief interaction and precipitation can all occur well ahead of a WD. By contrast, one day after the wind peak, as the WD approaches the UIB, the zonal gradient of geopotential at 700 hPa disappears, stopping the cross-barrier meridional flow (Figure 4.3-C and F). The subsequent intrusion of colder air below the WD is then responsible for ending the interaction of the WD and the relief.

4.4.2 PC Regressions

As mentioned in Section 4.1, a systematic quantification of the link between winds triggering precipitation and WDs is needed. A PC regression is used again, but with different variables. The predictand (Y) is $W700$, as computed in Section 4.3. Both positive and negative values of $W700$ are considered despite the fact that only positive values contribute to precipitation; this choice will be discussed when addressing the results in the paragraphs below. The predictors are the 3-hourly time series of geopotential heights at 300 hPa. The use of the geopotential height to characterise WDs is justified by the pattern seen in Figure 4.2 as well as by past tracking analysis (Cannon et al., 2016a) and WD indices (Madhura et al., 2015; Midhuna et al., 2020). Time series of geopotential heights are selected for each grid point separated by 1° and within a box between 50° and 90°E, and between 20° and 50°N as indicated by the domain shown in Figure 4.4-A. Higher resolution is not needed and requires significantly higher computational power. The time series are standardised (i.e. the mean is removed and the result divided by the standard deviation) before the decomposition into principal components. The first 30 principal components are selected for the regression. The coefficients of the regression are displayed in Figure 4.4-A and the scatter plot of $W700$ against its prediction is shown in Figure 4.4-B.

Figure 4.4-A shows an area of negative values (dark blue colours), which is much smaller than the extent of negative geopotential anomaly shown in Figure 4.2-B and centred slightly eastward. This area is surrounded by a ring of positive values. The gradient formed indicates a cyclonic anomaly[1]. The ring can be interpreted as the effective radius of WDs in triggering southerly winds in the UIB. More positive values are located over the north of the Tibetan Plateau indicating a stronger importance of geopotential gradient to the east of the disturbance. This pattern confirms the importance of the area of positive geopotential anomaly to the east of the main feature of the WD, discussed above (Figure 4.2-A and B).

The R^2 of the regression is 0.668, and Figure 4.4-B confirms the general good agreement between predictand and prediction. However, the fit worsens for values of $W700$ above $2.5\,\mathrm{m\,s^{-1}}$: these values are underestimated by the prediction, either suggesting the lack of other predictors or a non-linear relationship between $W700$ and WD intensity. Including negative values of $W700$ does not affect the pattern in Figure 4.4-A but is important to demonstrate the linear relationship for values between -2 and $2\,\mathrm{m\,s^{-1}}$. Other predictors are explored to improve the prediction for high $W700$.

[1]It could also be an anticyclonic anomaly, the prediction will then be negative, corresponding to northerly winds in the UIB.

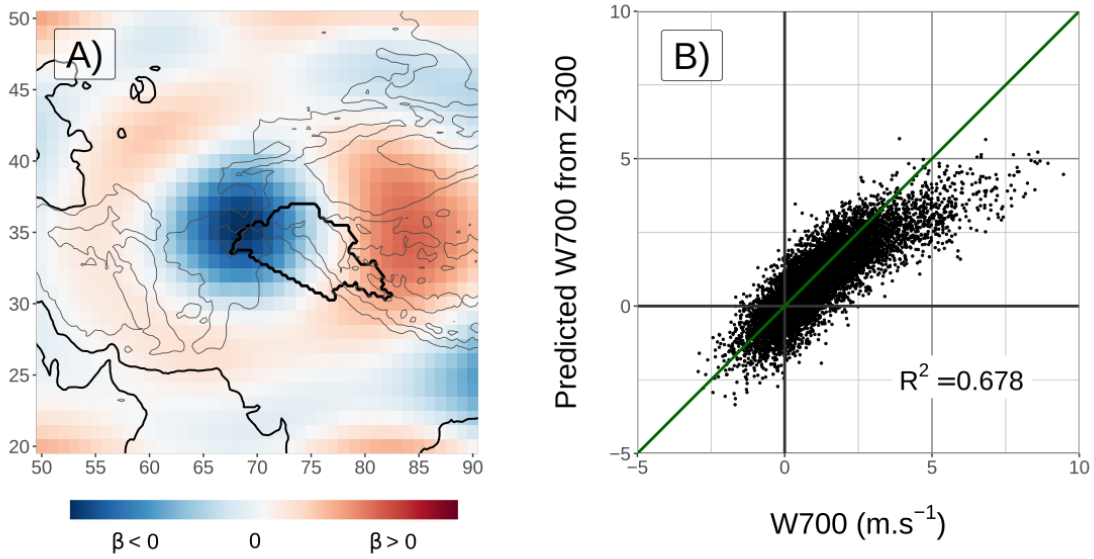


Fig. 4.4 Results of the regression of 3-hourly $W700$ with geopotential height at 300 hPa ($Z300$). Panel A shows the coefficients β associated with each grid point. The exact values of the coefficients are not interpretable and are therefore not indicated. Panel B is the scatterplot, with the green line representing the fit of the regression. The coefficient of determination (R^2) is also indicated.

Relative vorticity, which has been used for WD tracking by (Hunt et al., 2018a), produces a slightly lower R^2. Thus, vorticity is not as good a predictor as geopotential in our case. Regressions using geopotential height at lower level have a higher R^2, up to 0.878 for 700 hPa, the altitude where moisture transport is considered for the regression of precipitation. However, 700 hPa is less relevant for WDs, and the ring pattern displayed in Figure 4.4-A is replaced by a north-west/south-east dipole that is less characteristic of a WD and rather shows the relation with the geostrophic wind (not shown).

To connect the WD characteristics at 300 hPa with the geostrophic effect at 700 hPa, the regression is performed using the geopotential thickness between 300 hPa and 700 hPa, in addition to the geopotential at 300 hPa. The rationale behind this selection is that, under geostrophic equilibrium, the wind at 700 hPa can be related to geopotential at 300 hPa (or the geostrophic wind at that altitude), and geopotential thickness between those two altitudes (from which the thermal wind derives). In addition, the geopotential thickness is split in two layers at 500 hPa, characterising the upper and the lower troposphere respectively, so that geopotential thicknesses above the Tibetan Plateau can be considered. The first 40 principal components are selected for each of the three predictors. The results are presented in Figure 4.5.

Including geopotential thicknesses significantly increases the R^2 of the regression from 0.668 to 0.882. The scatter plot on Figure 4.5-A shows that the underestimation of high $W700$ values has mostly been corrected. Some underestimations remain for the highest values of $W700$ possibly because the larger latent heat release and increased buoyancy in the context of stronger convergence further sustain cross-barrier winds before the pressure gradient can adjust to it. Figure 4.5-B, C and D shows the pattern of the coefficients associated with the three predictors. Each of them exhibits a similar dipole feature across the UIB, although with an opposite sign for the geopotential thicknesses compared to the geopotential height. Indeed, a negative geopotential anomaly at 700 hPa can be explained by either a negative geopotential anomaly at 300 hPa or a positive geopotential thickness anomaly between the two altitudes. A ring feature is still evident for the geopotential at 300 hPa (Figure 4.5-B), whose centre is located to the west of the study area (Figure 4.5-B), where the relief forms a notch important to trigger the vertical velocities (cf. Chapter 3; also Baudouin et al., 2020a; Cannon et al., 2015; Lang and Barros, 2004). The displacement of the centre to the south-east compared to Figure 4.4-A is consistent with the tilt of the WDs with height (Figure 4.2-B and Figure 4.3-B; also Dimri and Chevuturi, 2014; Hunt et al., 2018a). Yet, the main feature, the dipole, relates to a south-westerly geostrophic wind. Accordingly, the dipole for the geopotential thicknesses (Figure 4.5-C and D) implies an increase in south-westerly winds towards the surface. Those patterns suggest a geostrophic wind that is rotated by approximately 45°

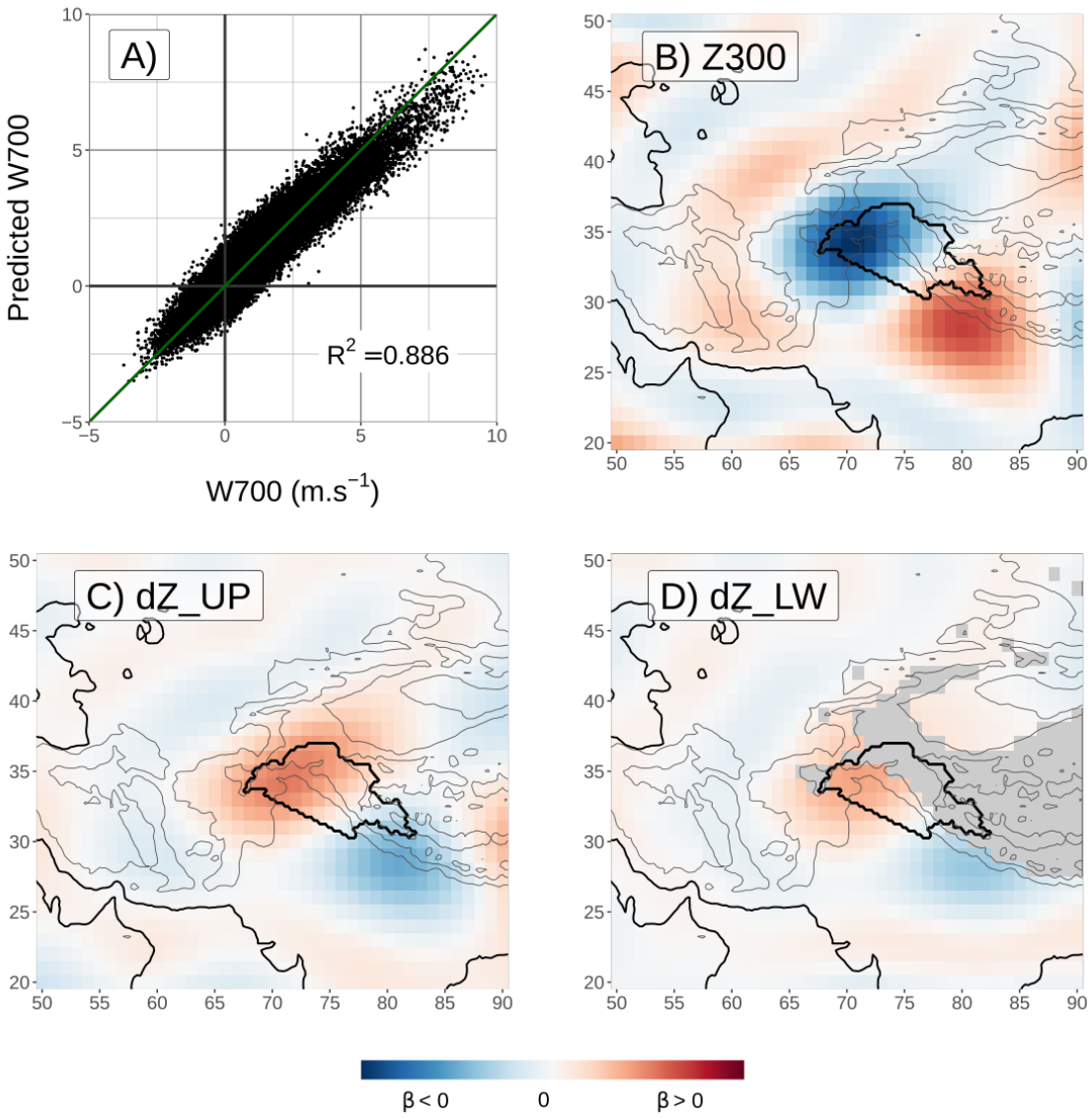


Fig. 4.5 Results of the regression of 3-hourly $W700$ with geopotential height at 300 hPa ($Z300$), geopotential thickness 500-300 hPa (dZ_UP), and geopotential thickness 700-500 hPa (dZ_LW). Panel A is the scatterplot as in Figure 4.4-B. Panels B, C, and D show the coefficients β associated with each predictors. The exact values of the coefficients are not interpretable and are therefore not indicated.

compared to the southerly winds that it is meant to predict (cf. Figure 4.1). Although the geopotential gradient is sufficient to predict wind intensity, it does not indicate the wind direction, which can be bent towards the lower pressures, as noted in Chapter 3 (Section 3.4.3.

The partial prediction of $W700$ associated with each of the three fields are computed using equation 3.7, and named $Z300$ for the geopotential at 300 hPa, and dZ_UP and dZ_LW for the geopotential thickness between the layers 500-300 hPa and 700-500 hPa respectively. Interestingly, both dZ_UP and dZ_LW are negatively correlated with $Z300$. This matches with the opposite pattern of regression coefficients (Figure 4.5-B, C, and D). In fact, dZ_UP and dZ_LW are even negatively correlated with the predictand, $W700$. In regression studies, predictors exhibiting such a behaviour are referred to as a suppressor: the predictor suppresses the variability of one or several other predictors (geopotential at 300 hPa in this case, cf. Nathans et al., 2012; Smith et al., 1992). To simplify the analysis, a new contribution of the geopotential thickness to the wind is computed. dZa_UP is the residual of the regression of dZ_UP with $Z300$: the variability of $Z300$ is removed from dZ_UP. It describes the anomalous geopotential thickness given the situation of the geopotential at 300 hPa. The same is performed for dZ_LW. The new contributions characterise the thermal structure of the WDs. If dZa_UP is positive (negative) then the northwest-southeast temperature gradient over the UIB is weaker (stronger) than usual. This will be further explored in the next section. For now, the regression can be summarised in the equation below, with $\widehat{W700}$ as the prediction of $W700$:

$$\begin{aligned}\widehat{W700} &= \beta_0 + Z300 + dZ\_UP + dZ\_LW \\ &= \beta_0 + Z300 + (a \times Z300 + dZa\_UP) + (b \times Z300 + dZa\_LW) \\ &= \beta_0 + (1 + a + b) \times Z300 + dZa\_UP + dZa\_LW\end{aligned} \tag{4.1}$$

Note that $Z300$ is recomputed hereafter so that it includes the fractions a and b derived from dZ_UP and dZ_LW, respectively. Note also the presence of the constant offset β_0.[2]

4.4.3 Thermal structure

dZa_UP and dZa_LW characterise the thermal structure of the atmosphere over the UIB in relation to the approach of a WD. As noted before, when $W700$ is high, a horizontal temperature gradient is present over the UIB (Figure 4.2-B), from the north-west (cold) to the south-east (warm). A positive dZa_UP means that the horizontal temperature gradient is

[2]There is no need of coefficients in front of the variables $Z300$, dZ_UP, or dZ_LW, as the regression coefficients are already included in the definition of these variables.

weaker than usual, indicating that the cold core of the WD is smoother, weaker. In contrast, a negative dZa_UP means that the horizontal temperature gradient is stronger and the cold core of the WD is deeper.

The different contributions to $W700$ change during a wind peak event as it is evident in Figure 4.6. In this figure, the peak of $W700$ is computed as a composite of time steps where $W700$ is above the 90^{th} percentile, as for Figure 4.2 and 4.3. The high value of $W700$ is first due to dZa_UP (accounting for 46% of the predicted value), highlighting the importance of considering the WD thermal structure. $Z300$ is close behind (43%), while dZa_LW has a minor role (11%). Interestingly, $W700$ (the dashed orange line in Figure 4.6) raises and decays symmetrically around its peak. However, this is not the case for its different contributions (other dashed lines). All contributions rise steadily before the wind peak. However, the two geopotential thickness contributions peak before $W700$ reaches its maximum intensity and quickly drop after: $W700$ is only maintained by $Z300$ after its peak. The asymmetry between $Z300$ and dZa_UP or dZa_LW can also be regarded as the result of the tilt or baroclinicity of the wave, as is evident from the comparison of these results with Figures 4.2 and 4.3. Focusing on Figure 4.2-A, two days before the wind peak, a gradient in geopotential is already present over the UIB in relation to the positive geopotential anomaly ahead of the WD. However, the warm anomaly of southerly advection implied by the configuration of the geopotential height anomaly limits the presence of any gradient in temperature (and thus of geopotential thickness). At the wind peak (Figure 4.2-B), the approaching WD tightens the gradient of the 300hPa geopotential while the colder air remains at the edge of the IUB, resulting in the tilting of the geopotential minimum with height. By contrast, two days after the wind peak, the WD enters the UIB with its cold core and the temperature gradient matches and counterbalances the geopotential gradient, so that southerly winds at 700 hPa weaken (Figure 4.2-C).

There exists variability in the thermal structure of the WD. To investigate this, a more complex composite analysis is used, based on a quantile regression. Two subsamples of the time steps with $W700$ above the 90^{th} percentile are created, so that they both have the same mean geopotential anomaly, but one is related to lower values of $W700$ (although still above the 90^{th} percentile) and the other to higher values. Two quantile regressions are used, each predicting the first and the third quartile of $\widehat{W700}$ (the prediction of $W700$ with geopotential heights and thicknesses, cf equation 4.1), as a function of $Z300$, of geopotential anomaly at 300 hPa at the centre of the mean WD (66°E - 36°N), and of months. The group with lower (higher) $W700$ is composed of all the time steps with $W700$ below (above) the first (third) quartile computed by the regressions. Since $W700$ is comparatively higher than $Z300$ before the wind peak than after (cf. Figure 4.6), using only $Z300$ in the quantile regressions would

lead to an over representation of time steps with WDs closer to (further away of) the UIB in the low (high) wind group. Including the geopotential anomaly at 300 hPa at the centre of the mean WD removes this issue and insures that the mean WD intensity is similar in both groups. Finally, the months are introduced in the quantile regressions, so that the composites of the anomaly presented below are not affected by the seasonality: for each month, the same number of time steps is present in each group. Note that the quartiles of $\widehat{W700}$ are predicted and not that of $W700$ itself, so that the differences between the two groups are not related to the variability missed by the PC regression of $W700$ with geopotential heights and thicknesses.

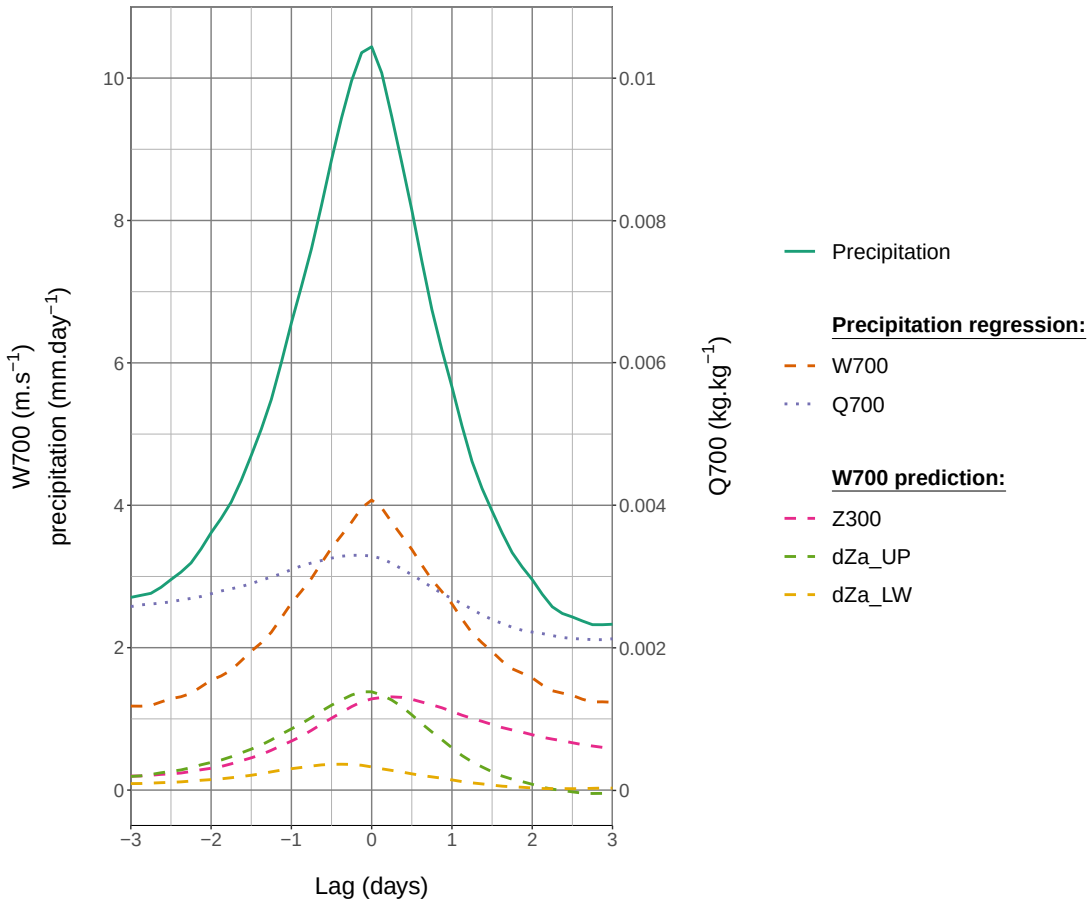


Fig. 4.6 Lead/lag composite around the 10% highest values of $W700$, for ERA5 precipitation, contributions to precipitation based on moisture transport ($W700$, $Q700$, cf. Section 4.3), and partial prediction of $W700$ based on geopotential ($Z300$, dZa_UP, and dZa_LW, cf. Section 4.4.2). Note that $Z300$, dZa_UP, and dZa_LW have the same units as $W700$, and, with the intercept, add up to the prediction of $W700$ ($\widehat{W700}$, cf. Equation 4.1).

Table 4.1 Mean values computed for higher $W700$ and lower $W700$ as defined in Section 4.4.3 for different variables: averaged precipitation in the UIB from ERA5; wind ($W700$) and moisture ($Q700$) contributions to precipitation, derived from the regression of precipitation with moisture transport (cf. Section 4.3); the prediction of $W700$ with geopotential height and thickness ($\widehat{W700}$), cf. Section 4.4.2). Finally the three last quantities are the partial prediction associated with $\widehat{W700}$ (cf. Equation 4.1). $Z300$ is the component associated with geopotential at 300 hPa, *dZa_UP* (*dZa_LW*) is the geopotential thickness anomaly for the layer 500-300 hPa (700-500 hPa) where the anomaly is relative to $Z300$. The sum of these three quantities with the offset ($0.7\,\mathrm{m\,s^{-1}}$) equal $\widehat{W700}$.

	Precipitation ($\mathrm{mm\cdot 3hrs^{-1}}$)	$W700$ ($\mathrm{m\,s^{-1}}$)	$Q700$ ($\mathrm{kg\,kg^{-1}}$)	$\widehat{W700}$ ($\mathrm{m\,s^{-1}}$)	$Z300$ ($\mathrm{m\,s^{-1}}$)	*dZa_UP* ($\mathrm{m\,s^{-1}}$)	*dZa_LW* ($\mathrm{m\,s^{-1}}$)
Lower $W700$	5.9	3.2	3×10^{-3}	2.6	1.3	0.4	0.2
Higher $W700$	16.8	5.3	3.6×10^{-3}	4.9	1.3	2.4	0.5

Table 4.1 presents the mean values of different quantities for the two groups. As per definition, $Z300$ has the same value in both groups, indicating that the geopotential height gradient across the UIB is similar. Yet, $W700$ is two thirds higher in the group with higher $W700$ than in the other group. The prediction of $W700$ ($\widehat{W700}$) is lower than $W700$ itself, due to the small underestimation noted in Figure 4.4-A, but the underestimation is similar in both cases (about $0.5\,\mathrm{m\,s^{-1}}$). Therefore, the difference in $W700$ between the two groups is only explained by the difference in geopotential thicknesses, and particularly *dZa_UP*.

Figures 4.7 and 4.8 represent the same variables as Figures 4.2 and 4.3, but for the group with lower $W700$ in the first row, and the one with higher $W700$ in the second row. As expected, the WD position and intensity are about the same when no lag is considered (Figure 4.7-B and E and Figure 4.8-A and C) in both groups, yet other characteristics of the WD change.

In the case of lower $W700$, the positive anomaly of geopotential height over the Tibetan Plateau (cf. Figure 4.7-B) is barely distinguishable, leading to smaller geopotential height anomaly gradients, despite having fixed $Z300$. Furthermore, the WD centre is co-located with a more intense cold core: the temperature gradient anomaly is parallel to that of the geopotential height, which contributes to reducing the cyclonic anomaly toward the ground (cf. *dZa_UP* close to 0 in Table 4.1). Both features induce a reduction of the cyclonic circulation at 700 hPa (Figure 4.8-A) and of southerly winds over the Indus Plain (Figure 4.8-B) and therefore of $W700$. The more symmetric cold core below the WD also induces

a smaller tilt, both characterising a weaker baroclinicity that is likely to contribute to the weaker low-level activity.

By contrast, when $W700$ is higher, a high positive anomaly of geopotential height is present over most of the Tibetan Plateau: a stronger east-west gradient of geopotential height is present over the UIB. The geopotential height gradient is at an angle compared to the pattern in Figure 4.4-B, so that $Z300$ can remain the same despite the tighter gradient. The temperature anomalies then contribute to reorientating the lower-level geopotential gradient in the observed south-east/north-west direction. Since the temperature anomaly is mostly parallel to the Himalayas (Figure 4.7-B, resulting in a higher dZa_UP in Table 4.1), zonal geopotential height gradient changes only weakly with the altitude (Figure 4.8-

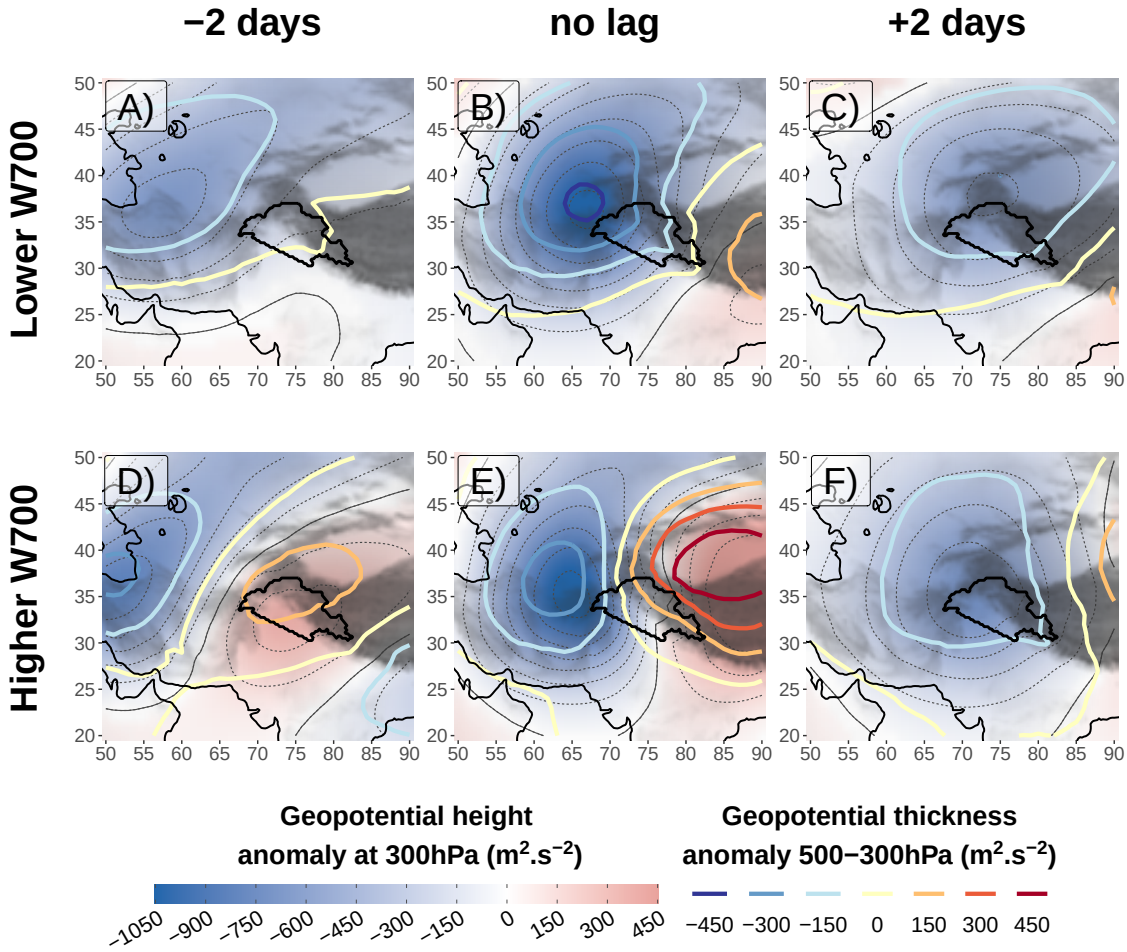


Fig. 4.7 Same as Figure 4.2, but the composites are based on a subsampling of lower (higher) values of $W700$, while still above the 90th percentile in panels A to C (D to F). The subsampling is made so that it is not impacted by seasonality nor by the intensity of the WD. This way, comparisons can be made with Figure 4.2.

C). Consequently, a stronger cyclonic circulation is present at 700 hPa, inducing stronger southerly winds over the Indus Plain (Figure 4.8-D) and thus cross-barrier winds in the UIB (cf. Higher $W700$ in Table 4.1). The low-level circulation propagates from the upper-level WD with a tilt (Figure 4.8-C), which, alongside the displacement of the cold core to the north-west of the WD centre (Figure 4.4-E), indicates a stronger baroclinicity, one of the causes for a more active dynamic.

Furthermore, a positive feedback mechanism occurs between the altitude temperature gradient and the low-level winds. The up-lift over the UIB mixes the air mass aloft and warms

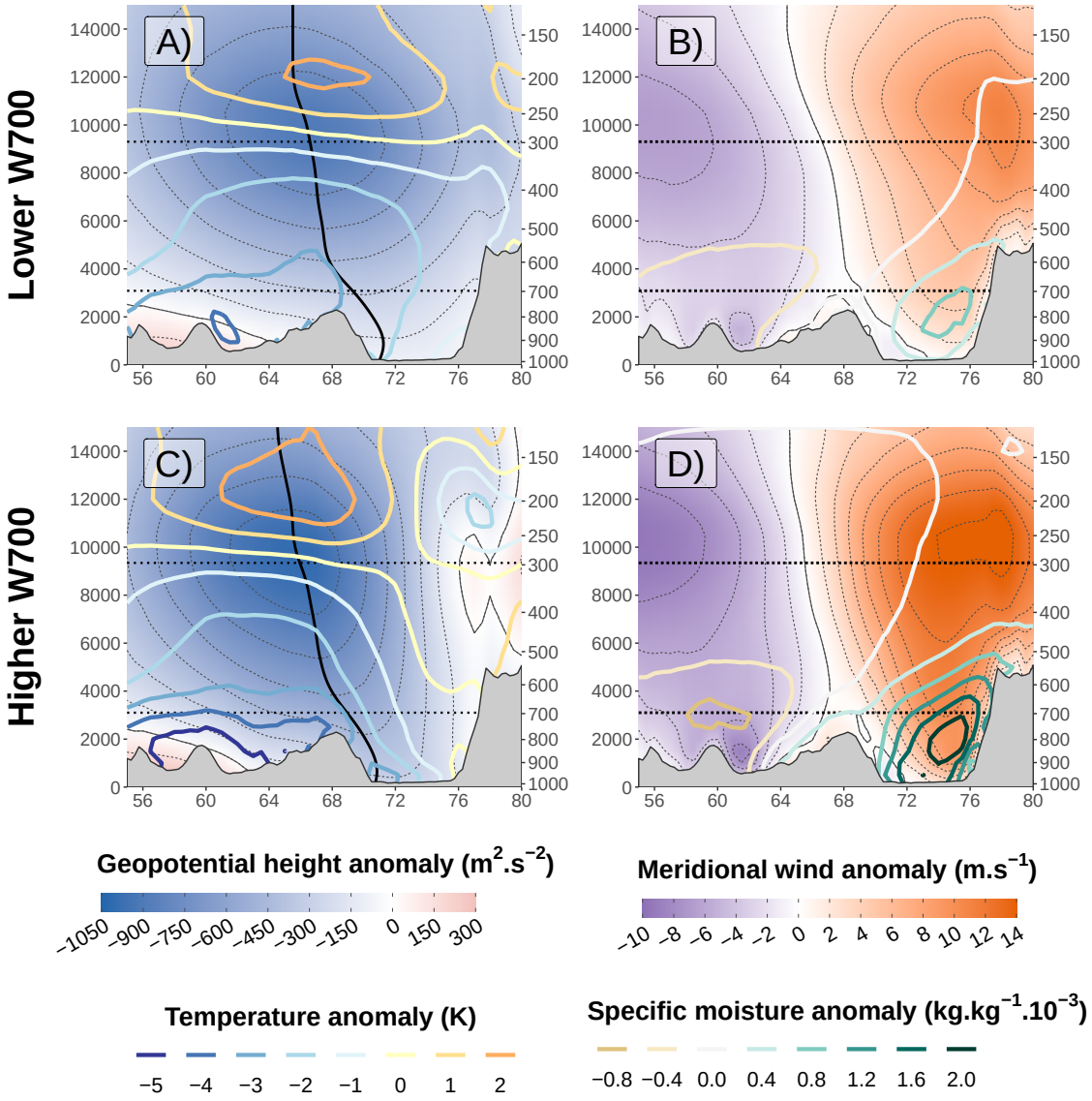


Fig. 4.8 Same as Figure 4.3 but without lag, and for composites based on the same selection as in Figure 4.7.

it through advection and latent heat release, which prevents the cold air from extending into the UIB alongside with the WD (Figure 4.8-C, 76°E, Figure 4.7-E). The zonal temperature gradient at altitude is therefore reduced, which in turn increases the geopotential height gradient at lower-level and thus the cross-barrier winds and the uplift. This effect can also explain the underestimation of $W700$ when only $Z300$ is considered (Figure 4.4-B) as well as the general potential for intense cross-barrier winds in the area.

4.5 Relating wind contribution and WDs to the larger scale atmospheric circulation

4.5.1 Interaction with the Subtropical Westerly Jet

The upper troposphere disturbance characterising a WD is embedded in the Subtropical Westerly Jet (SWJ Dimri et al., 2015; Hunt et al., 2018a). The SWJ circles the North Hemisphere in winter (Schiemann et al., 2009), and is related to the poleward motion of air masses at the top of the Hadley circulation (Krishnamurti, 1961). The SWJ also coincides with an area of stronger temperature gradient below it, through the thermal wind equilibrium (Schiemann et al., 2009). In Asia, the SWJ oscillates between two stable states: one north of the Tibetan Plateau, always present in summer, and one south of it, reached only in winter (Schiemann et al., 2009). The Tibetan Plateau also disrupts the horizontal propagation of the SWJ, which is split in two climatological jet streaks of higher intensity (Krishnamurti, 1961; Schiemann et al., 2009): one over the Arabian Peninsula (Arabian Jet, de Vries et al., 2016; Yang et al., 2004), and the other over East Asia (East Asian Jet, Xueyuan and Yaocun, 2005).

It has been argued that the position and intensity of the SWJ influences WD intensity at the intra-seasonal and inter-annual scale (Ahmed et al., 2019; Dimri et al., 2015; Filippi et al., 2014; Hunt et al., 2018a). These authors suggested that the higher kinetic energy in the SWJ is able to fuel the development of the WDs as for other baroclinic waves (e.g. Hunt et al., 2018a). Yet, Hunt et al. (2018a) mentioned that WDs are immature baroclinic waves that differ from their mature counter-part in the Atlantic or Pacific Ocean. There is a need for a better characterisation of the coupling between SWJ and WDs at the synoptic scale. Furthermore, little is known about the relation between the jet and the vertical velocities induced by the WD interaction with the relief, despite the importance for precipitation generation. This section describes the upper-troposphere large scale circulation, including the SWJ, during the passing of a WD.

A composite of wind speed and wind speed anomaly at 250 hPa as WDs pass is presented in Figure 4.9. As for previous figures (4.2 and 4.3) the composites are defined as the time

steps with $W700$ above its 90th percentile. The altitude of wind speed chosen is the one of maximum meridional wind (Figures 4.3-C and D and 4.8-B and D) close to the level where the jet is maximum (Hunt et al., 2018a; Midhuna et al., 2020; Schiemann et al., 2009; Xueyuan and Yaocun, 2005). The anomaly relates to the difference with the climatological wind speed, taking into account the seasonality of the time steps selected. In each of the four lagged composites, the SWJ is clearly present between 25°N to the western edge of the Figure 4.9, to 35°N to its eastern edge, with the two maxima mentioned earlier: the strongest south of Japan (East Asian Jet, above 55 $\mathrm{m\,s^{-1}}$), the other between Egypt and Iran (Arabian Jet, above 45 $\mathrm{m\,s^{-1}}$).

All lagged composites in Figure 4.9 show a north-south dipole of wind speed anomaly that progresses eastward, following the WD (cf. Figure 4.2). The dipole marks the presence of the anomalous cyclonic circulation and results in the narrowing and strengthening of the Arabian Jet. Dimri et al. (2015) suggests that part of the strengthening could be due to the merging of the SWJ with the polar jet, which marks the limit between cold and warm air at the surface. This is not evident from the figure, but the composite analysis may hide this dynamic due to the highly variable position of the polar jet. After the peak of $W700$ and the interaction of the WD with the relief, the two parts of the dipole dissociate themselves: the negative northern anomaly slowly moves north-eastwards, while the positive anomaly quickly continues eastward, south of the Tibetan Plateau, carried by the SWJ which exhibits then a third maxima (Figure 4.9-D).

When the lag is negative (Figure 4.9-A and B), the positive anomaly of wind speed further extends to the north-east, towards Central Asia, outside of the jet core. This deformation of the jet could be an effect of the interaction with the relief, whereby the SWJ tends to reach its second stable position, north of the Tibetan Plateau, which sometimes occurs in winter (jet split, Schiemann et al., 2009). A picture from Pisharoty and Desai (1956), still in use in Dimri et al. (2015), suggests on the contrary that stronger wind speed occurs at altitude at the rear of a WD. While some WDs may have this characteristic, WDs triggering high $W700$ and thus high precipitation do not exhibit that feature.

As the WD approaches the Tibetan Plateau, a maximum anomaly of wind speed becomes evident over the high grounds: first over the Pamir range (Figure 4.9-B) and then along the Himalayas (Figure 4.9-C and D). This anomaly is the result of the increased funnelling of the jet over the mountains. In the meantime, the WD starts deforming the positive anomaly of wind speed, eventually splitting it in half (Figure 4.9-B and C). The increased wind speed anomaly to the west of the WD is also likely the result of the interaction with the relief: the low-level convergence in the UIB sends air masses to the upper troposphere, which increases wind speed, forming an outflow. The outflow is characterised by a swift anticyclonic turn,

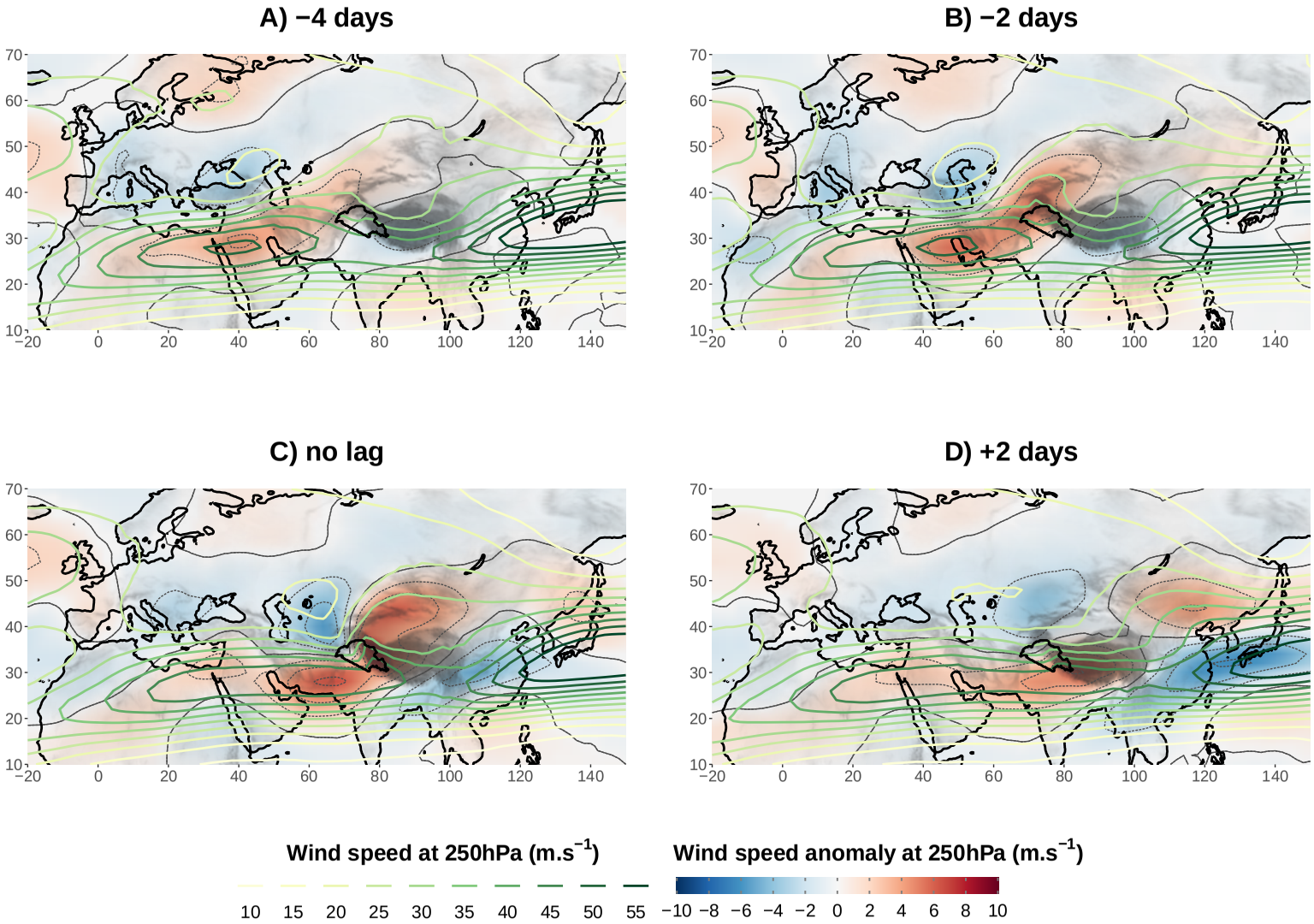


Fig. 4.9 Lead/lag composite maps of wind speed (thick contour lines) and wind speed anomaly (colour shading, thin contour lines every 2 m s^{-1}) at 250 hPa, based on the same selection of the 10% highest $W700$ as in Figure 4.2. Non-significant anomalies at the level 95% are shown in white.

with strong similarity to the warm conveyor belt associated with mature baroclinic waves (Martínez-Alvarado et al., 2014). The increase in latitude and the latent heat release are both key to explaining that change in relative vorticity (Grams et al., 2011). The outflow can also be revealed by the analysis of the cloud cover and the development of large bank of cirrus (Agnihotri and Singh, 1982; Hunt et al., 2018a; Rakesh et al., 2009).

The characteristics of this outflow suggest that the uplift provides a baroclinic feedback effect on the WD intensity during the interaction of the WD with the relief (cf. also WD slowing and growing process discussed in Section 4.4.3). The baroclinic effects can be described using potential vorticity (PV), as the WD is characterised by a high PV anomaly (Hunt et al., 2018a). When a negative PV anomaly finds itself in a shear environment, baroclinically forced vertical velocities occur: with uplift upwind the anomaly, and subsidence downwind (Malardel, 2005). This pattern is indeed what happens for a WD when embedded in the SWJ. More specifically, it is the zonal gradient of PV ahead of the WD that triggers the baroclinic uplift. When a WD approaches the UIB, its interaction with the relief results in an orographic uplift, which adds up with the baroclinic uplift. However, the uplift also results in decreased PV at high altitude (300 hPa) over the UIB due to the advection of low-level lower PV as well as diabatic heating. Hence, the zonal PV gradient increases and stronger baroclinic uplift occurs, resulting in a positive feedback. However, once the high PV anomaly associated with the WD enters the UIB, the mechanism reverses: the orographic uplift results in the weakening of the PV anomaly, eventually leading to the decay of the WD. This interaction and the intense uplift is revealed by the deformation of the jet near the UIB during the passing of the WD (Figure 4.9). Particularly, when the uplift is maximum, the jet configuration is characteristic of a strong divergence at altitude over the IUB: the study area is located near the eastern exit of the Arabian Jet, and in the right entrance of the outflow[3]. This mechanism is similar to the baroclinic interaction between a tropopause anomaly and a surface disturbance, but where the surface disturbance is replaced by the relief. It also results in a weaker, shorter-lived interaction.

Finally, after the peak of $W700$, the positive anomaly of wind speed related to the outflow breaks free from the circulation associated with the WD and continues eastward, north of the East Asian Jet. By contrast, the WD-relief interaction contributes to lower the intensity of the SWJ downwind: the East Asian Jet is notably weakened as a result (Figure 4.9-D). Meanwhile, the Arabian Jet remains anomalously strong; in fact, two days after the peak of $W700$ the strength of the anomaly is about the same as four days before the peak. Hence, the increased intensity of the Arabian Jet is a longer-term feature that seems to promote the

[3]On a jet streak, the right entrance and the left exit are characterised by baroclinic upper-level convergence, (Malardel, 2005)

intensity or occurrence of a WD, but is not directly affected by the passing of one. This feature is evident from several intra-seasonal and inter-annual studies (Ahmed et al., 2019; Filippi et al., 2014; Hunt et al., 2018a). Hunt et al. (2018a) also discusses the SWJ position, but the synoptic analysis presented here does not indicate a change in the SWJ position.

4.5.2 Other large scale characteristics: Wave patterns and blocking events

It is evident that WDs are related to other circulation patterns that are not easily identified in the analysis of the jet in Figure 4.9. The composite analysis of geopotential height at 500 hPa over the same selection of time steps, presented in Figure 4.10, helps to distinguish the different centres of action and to characterise the wave pattern.

The WD is clearly visible on Figure 4.10-A, as a negative anomaly in geopotential height to the west of the UIB. The WD is part of a wave train (a succession of positive and negative anomalies) spreading to the west across an area stretching from North Africa to the UIB. The wave train indicates the relation of WDs with Rossby waves (Hunt et al., 2018b). Yet, the pattern is not very pronounced, suggesting significant case-to-case variability. The WD is also in phase with a second wave train over the Pacific Ocean. This wave pattern may be the continuation of the western wave train, although the Tibetan Plateau splits the two.

Beside the WD, the strongest anomaly in Figure 4.10-A is a positive anomaly located over western Russia whose shape is similar to an anticyclonic blocking. Considering the WD to the south-east, and another minimum to the south-west, it forms an omega shape, a typical sign of a blocking event (e.g. Woollings et al., 2018). Further, this positive anomaly is already present four days before the peak of $W700$, slightly more westward, and more pronounced (Figure 4.10-B). The pattern meets the requirement of duration of blocking events (Tibaldi and Molteni, 1990). However, the anomalies involved are not high enough to reverse the geopotential gradient, which is another characteristic of blocking (purple lines, Figure 4.10-A and B), although the averaging may have removed this characteristic.

One exceptional blocking event that occurred in western Russia in summer 2011 was related to extreme flooding in the Indus River basin (Galarneau et al., 2012; Lau and Kim, 2012; Yamada et al., 2016). This event did not happen during the WD active season, but demonstrates that blocking events do impact the weather in the UIB. Yet, blocking events have not been related to WDs in the literature before. The longitude of the anomalous pattern in Figure 4.10-B (30°E) is an area with high frequency of blocking occurrence (Barriopedro et al., 2006; Tibaldi and Molteni, 1990). These blockings generally occur at a higher latitude (around 70°N), however, and are known as Scandinavian blocking (Michel et al., 2012;

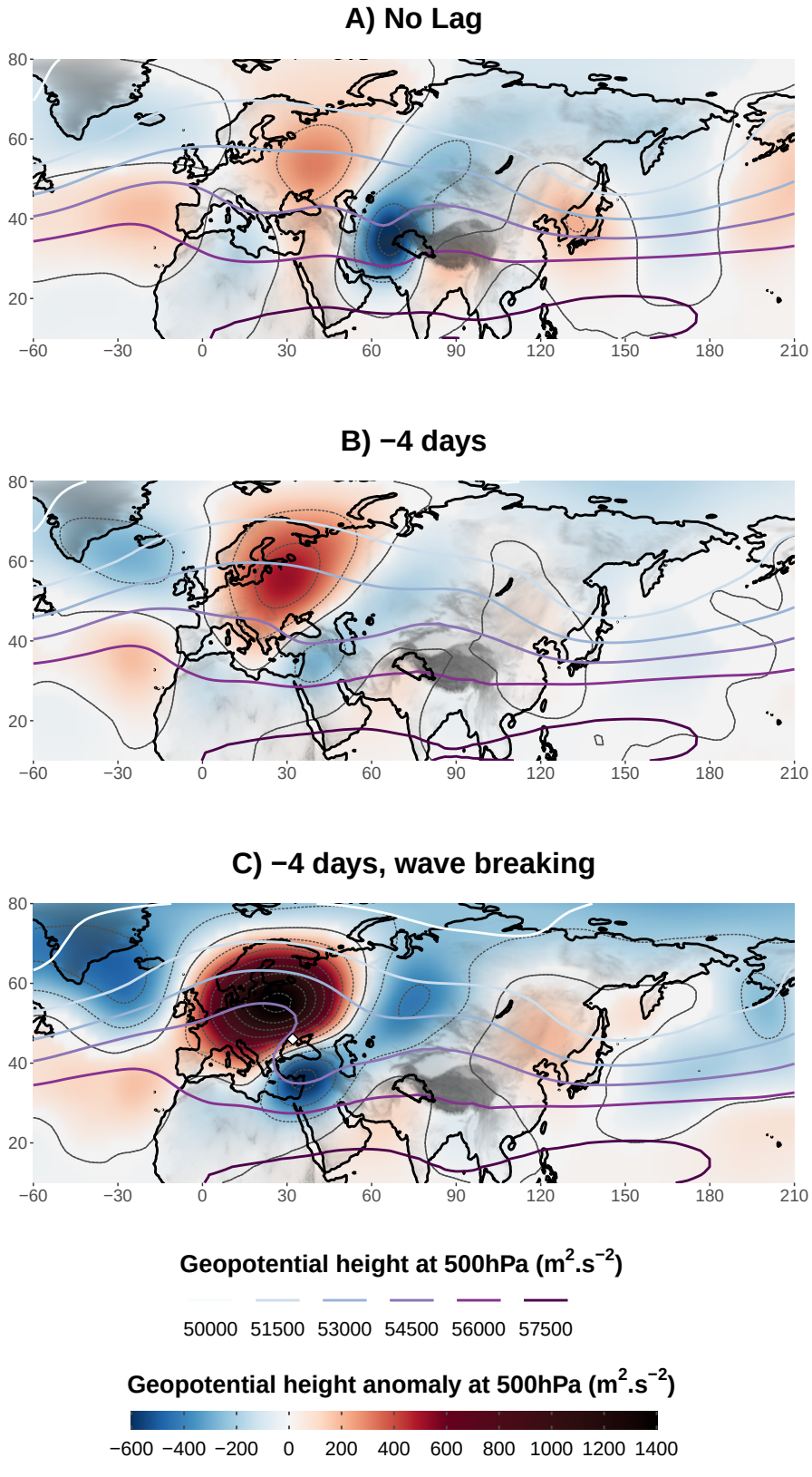

A) No Lag
B) −4 days
C) −4 days, wave breaking
80
60
40
20
−60
−30
0
30
60
90
120
150
180
210
Geopotential height at 500hPa (m².s⁻²)
50000 51500 53000 54500 56000 57500
Geopotential height anomaly at 500hPa (m².s⁻²)
−600 −400 −200 0 200 400 600 800 1000 1200 1400

Fig. 4.10 (Previous page) Composite maps of geopotential height (thick contour lines) and geopotential height anomaly (colour shading, thin contour lines every 200 $m^2 s^{-2}$) at 500 hPa. In panel A, the composite is based on the 10% highest $W700$ as in Figure 4.2, while a four days lead is used in panel B. For panel C, the selection also requires an easterly component of the wind at 250 hPa at 32°E and 42°N (white diamond) four days before the peak of $W700$ (subset of the selection for panel B). Non-significant anomalies at the level 95% are shown in white.

Scherrer et al., 2006); they form one of the four main weather regimes in Europe (Cassou et al., 2004). The southern position of the anomaly allows cold air advection toward the WD that can fuel its development, particularly as the WD moves eastward and aligns with a trough located to the east of the positive anomaly.

Figure 4.10-C shows the composite map of a sub-selection of the timesteps characterised by reversal of the circulation (easterly winds) north of the Black Sea, 4 days before (White diamond). It corresponds to about one third of the time steps selected in Figure 4.10-A or B (i.e. one thirtieth of the observations available). Under this condition, the anticyclonic anomaly encompasses most of Central and Eastern Europe while the precursor of the WD is located to its south, over the eastern Mediterranean Sea (Figure 4.10C). This configuration is similar to a Rex block (Rex, 1950). Here, however, due to the lower latitude of the anomaly extrema (below 60°N), the anomalous anticyclonic circulation does not split from the subtropical ridge, but rather extends from it across Europe (purple lines, Figure 4.10-C). The curvature of the 54 500 $m^2 s^{-2}$ contour line instead indicates the presence of a Rossby wave breaking. Wave breakings can initiate blocking events (Masato et al., 2012; Michel et al., 2012; Pelly and Hoskins, 2003), but they do not share the blocking's constraints on duration. Here, the equatorward motion to the south-east of the anomalous anticyclonic anomaly probably excites a wave in the SWJ that eventually propagates eastward and develops into a WD. Regardless, the persistence of the wave breaking does not seem to be relevant in the growing of a WD, thus suggesting that a fully established blocking event is not required. Interestingly, the selected time steps with zonal wind reversal at 250 hPa do not exhibit significant differences in terms of precipitation (not shown).

The last feature evident from all panels of Figure 4.10 is a dipole over the Atlantic. This dipole is characteristic of a positive phase of the North Atlantic Oscillation (NAO+) in which both Icelandic low and Azores High are stronger. Many studies have shown that the number of WD or the precipitation in the UIB are enhanced (lessened) during a positive (negative) phase of NAO, at a monthly time scale (Filippi et al., 2014; Kar and Rana, 2013; Syed et al., 2010, 2006; Yadav et al., 2009). A positive NAO results in an enhanced westerly flow over the

Atlantic and Europe. Particularly, the Atlantic Jet is significantly stronger, and shifts slightly south, as it can be seen in all panels of Figure 4.9 around 50-60°N. NAO+ is also related to a stronger Hadley circulation, which intensifies the SWJ and moves it to the south (Filippi et al., 2014). Filippi et al. (2014) and Hunt et al. (2018a) argue that these characteristics of the SWJ lead to more numerous or more intense WDs, therefore explaining the link between precipitation in the UIB and NAO. However, the synoptic analysis performed here suggests a more detailed mechanism. The stronger Atlantic Jet during NAO+ drives more energetic Rossby waves across Europe, leading to the formation of transient but intense anticyclonic anomalies over Eastern Europe (Strong and Magnusdottir (2008). These anomalies either directly trigger a WD via Rossby wave breaking, or simply fuel its development by directing cold air southward and increasing the Arabian Jet intensity as discussed above. Further, Filippi et al. (2014) showed that the meridional tilting of the jet modulates the strength of the relationship between precipitation in the UIB and NAO. This link is strengthened when the jet exhibits a more pronounced meridional tilting (to the north-west). Hence, a rather small meridional component of the jet is needed for the development of the anticyclonic anomaly over Eastern Europe.

4.5.3 Relation to the thermal structure

Depending on the thermal structure of the WD, the large-scale circulation is slightly different. The two composites defined for Figures 4.7 and 4.8 are reused here (same WD intensity but different $W700$). Figure 4.11 presents the mean wind speed for the two groups. In the case of a lower $W700$ (Figure 4.11-A), the wind anomaly of the outflow is weaker than expected. The SWJ is not weaker, however, but instead expands toward the south of the Tibetan Plateau. Consequently, the UIB is no longer in the area of enhanced upper level divergence of the jet. The lower baroclinicity in this group and the slower propagation are likely due to a weaker interaction between the WD and the SWJ (e.g. the wind shear and the cyclonic anomaly are further apart). Moreover, a second wind speed maximum is present at the same longitude, but is around 60°N. This anomaly may indicate the presence of a stronger polar jet, potentially drawing energy from the SWJ. The analysis of the composite of geopotential height at 500 hPa does not reveal the origin of the WD precursor (Figure 4.12-A), likely because the composite averages very different situations. Either the WDs have different origin, or they only intensify close to the UIB.

In the case of higher values of $W700$ (Figure 4.11-B), the situation is opposite, with stronger outflow and a weakened jet south of the Tibetan Plateau. The composite of geopotential height suggests the presence of wave breaking (purple line in Figure 4.12-B). The

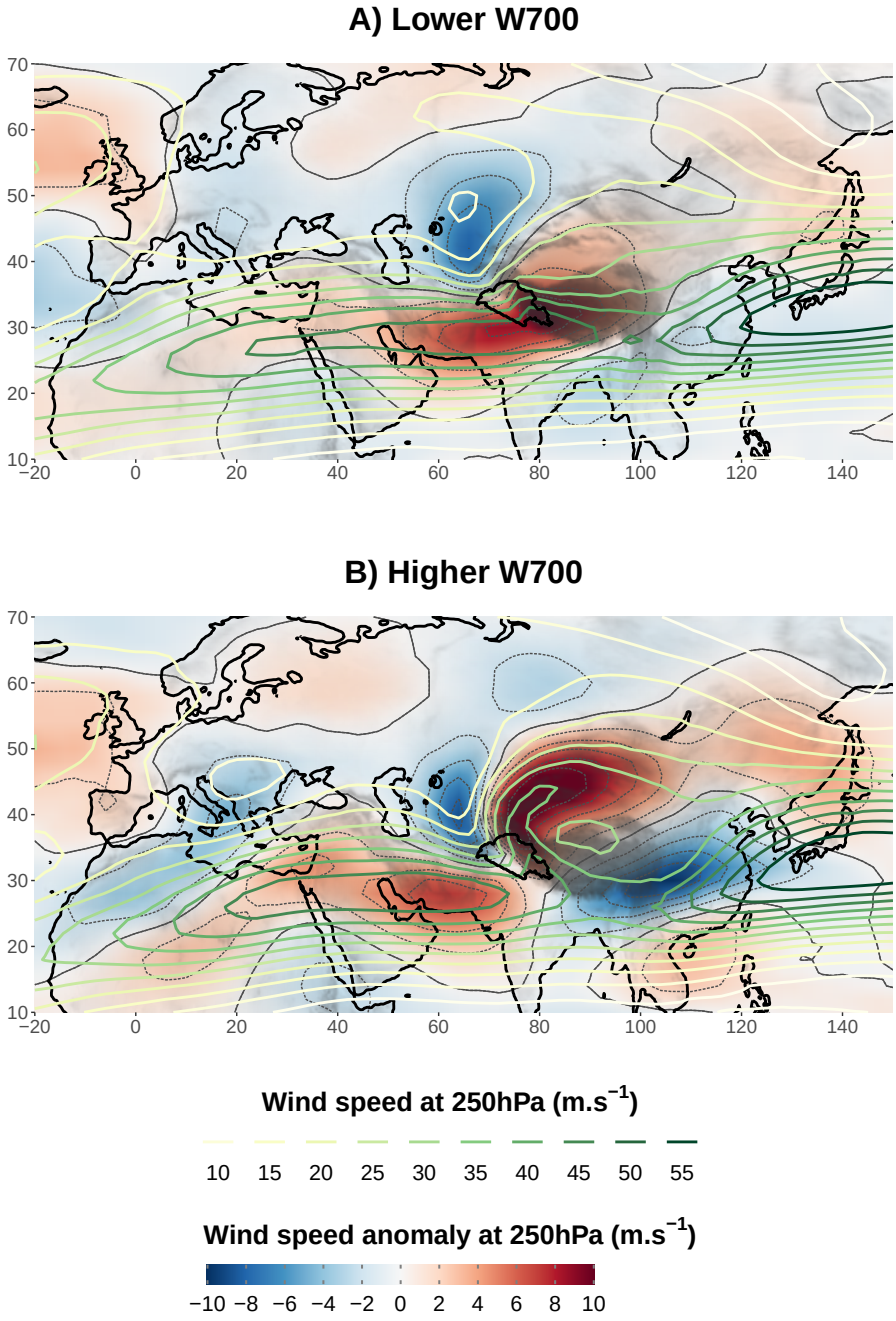


Fig. 4.11 Same as Figure 4.9-C (without lag), but for composites based on the same selection as in Figure 4.7.

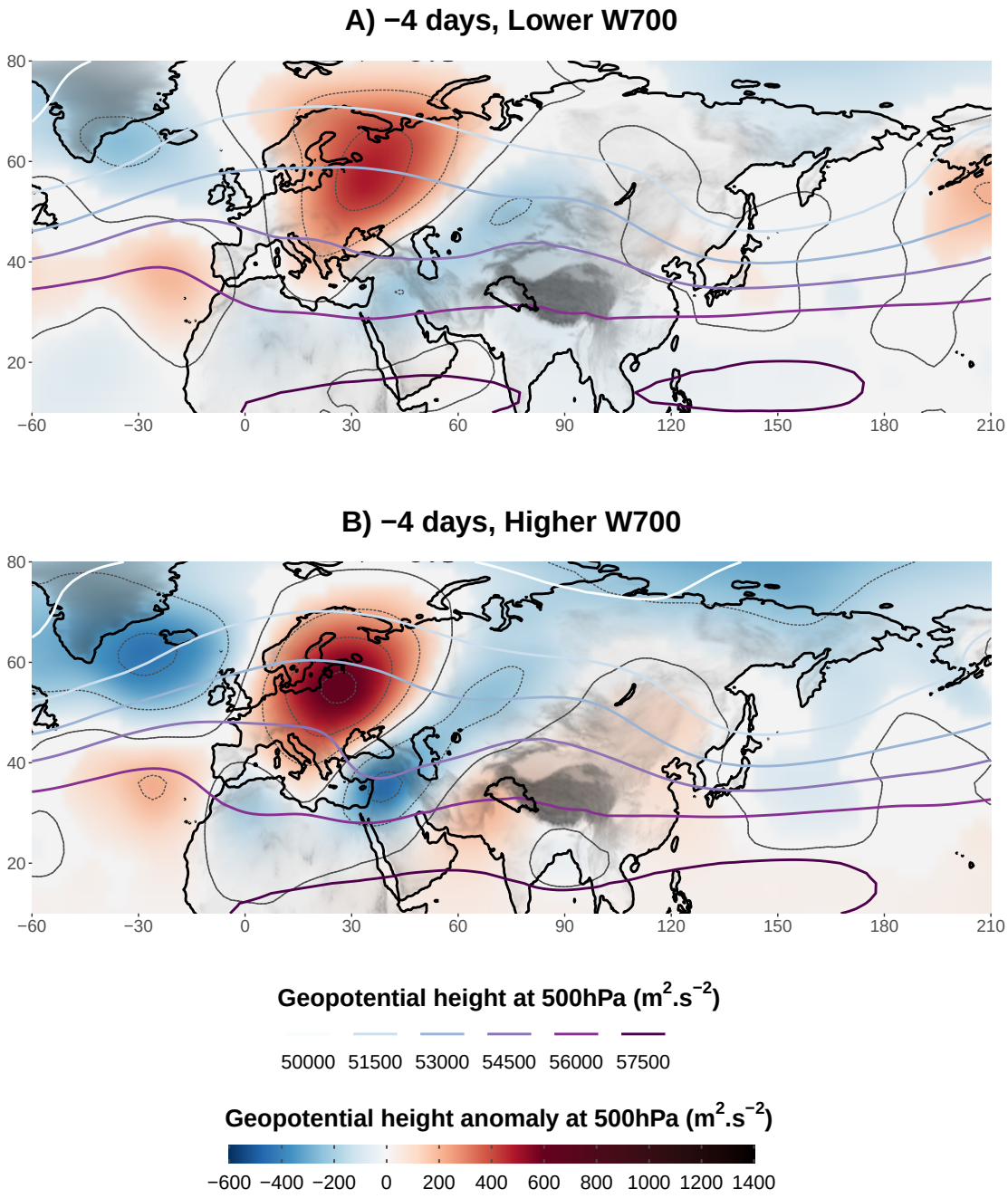


Fig. 4.12 Same as Figure 4.10-B (four days lead), but for composites based on the same selection as in Figure 4.7.

precursor of the WDs, when initiated by wave breaking, is further embedded in the SWJ and is therefore more strongly influenced by baroclinic processes.

To summarise, WDs develop on the left exit of an anomalously intense Arabian Jet, which fuels their growth though baroclinic processes. Their interaction with the relief produces an outflow north of the Tibetan Plateau, with effects downwind, particularly on the East Asian Jet, for the following days. These anomalies are proportional to the intensity of $W700$, independently from the WD intensity at altitude. Furthermore, the most intense WDs originate from the eastern Mediterranean Sea, initiated by a Rossby wave breaking over Europe. This mechanism helps to explain the relation between NAO and WDs discussed in the literature.

4.6 Relating moisture contribution to Western Disturbances

4.6.1 General

While wind is the most important parameter to explain precipitation synoptic variability, moisture content modulates the strength of the relationship (Chapter 3). Previous studies have investigated moisture transport in the context of winter precipitation in the UIB using moisture flux (Dimri, 2007; Filippi et al., 2014; Hunt et al., 2018a; Syed et al., 2010) or back trajectories (Boschi and Lucarini, 2019; Hunt et al., 2018b; Jeelani et al., 2018). Yet, some uncertainty remains about the moisture sources for precipitation and its pathways. The Arabian Sea is often suggested as the primary source of moisture (Dimri, 2007; Filippi et al., 2014; Hunt et al., 2018b). Uncertainty remains about the input of moisture from the Mediterranean Sea, as it is sometimes suggested that it is the origin of the most intense WDs that reach the UIB (e.g. Dimri et al., 2015; Filippi et al., 2014). Occasionally, the Red Sea (Dimri, 2007; Filippi et al., 2014), the Caspian Sea (Dimri and Niyogi, 2013; Syed et al., 2010) or even the Atlantic Ocean (Dimri et al., 2015) are also mentioned. The moisture pathway is certainly affected by the passing of a WD, through the deformation of the low-level wind field the WD imposes (Chapter 3). Yet, the reason for moisture variability at the synoptic scale has never been investigated properly. Consequently, this section focuses on the moisture contribution ($Q700$) as computed in Section 4.3. No evident variable, such as the geopotential height for the $W700$, can be used as a predictor of $Q700$ in a PC regression with good predictive skills. Therefore, the entire analysis of this chapter is based on composites. Further, since moisture matters for precipitation only when $W700$ is high enough, only the timesteps with $W700$ above the 90^{th} percentile are considered, as for previous composites (e.g. Figures 4.2 and 4.3).

4.6.2 Moisture pathway to the UIB

The isotopic ratio of hydrogen (δD) or oxygen (δ^{18}O) in rainwater can be used to understand the source of moisture and its transformation (Hussain et al., 2015b; Jeelani et al., 2018). Indeed, the isotopic ratio of water vapour is first set by the isotopic signature of the environment it evaporates from. Moisture is then depleted of its heaviest isotopes as it is transported further away from its source. However, the different potential sources of moisture mentioned in the literature (cf. previous paragraph) have all a similar isotopic signature, which limits the possibilities of this method (LeGrande and Schmidt, 2006).

Besides direct observations, numerical methods based on reanalysis data can be used to analyse moisture origin. High resolution modelling has made backtracking of air parcels possible in this context (e.g. quantitative Lagrangian approach). This tool is powerful for understanding the atmospheric circulation during specific events. However, it requires significant computing resources for a more exhaustive analysis like the one aimed for here. Various starting locations (longitude, latitude, altitude) of the air parcel that is backtracked would need to be considered as well as different time steps, leading to a more complex analysis. For comparison, almost 8000 time steps are considered at a 3-hourly frequency for the composite analysis (e.g. Figure 4.2), where Hunt et al. (2018b) backtracked air parcels for the 100 highest daily precipitation events and the latitude/longitude where those events occurred. The composite of moisture flux, coupled to that of precipitable water and evaporation, offers a simpler way to analyse the moisture movements, although it possibly oversimplifies the picture (qualitative Eulerian approach).

Figure 4.13 presents the composites of precipitable water (total column water) and vertical integral of water vapour flux (total column moisture transport). The seasonal mean (A) is weighted by the seasonality of occurrence of $W700$ above the 90$^{\text{th}}$ percentile. The panel shows that the UIB and the Indus Plain are dominated by a westerly moisture flux. This direction does not directly imply a westward origin of the moisture since precipitable water is over 15mm in the Indus Plain, while it is below 10 mm to the west of the plain, mostly due to the higher elevations (e.g. Suleiman range, Hindu Kush, Iranian plateau). Transient circulation (e.g. WDs) and local evaporation can explain this higher amount of precipitable water. In fact, the moisture content rises during a peak of $W700$ and in the days before as shown in Figures 4.3-C-D and 4.6, indicating the importance of WDs in moisture variability. Therefore, to understand the origin of moisture that triggers precipitation, moisture flux needs to be investigated during the passing of a WD, which is done using lagged composites in Figure 4.13-B to G.

Figure 4.13-E shows a significant accumulation of precipitable water in the UIB when $W700$ is maximum: the anomaly represents an increase of 30 to 50% of the precipitable

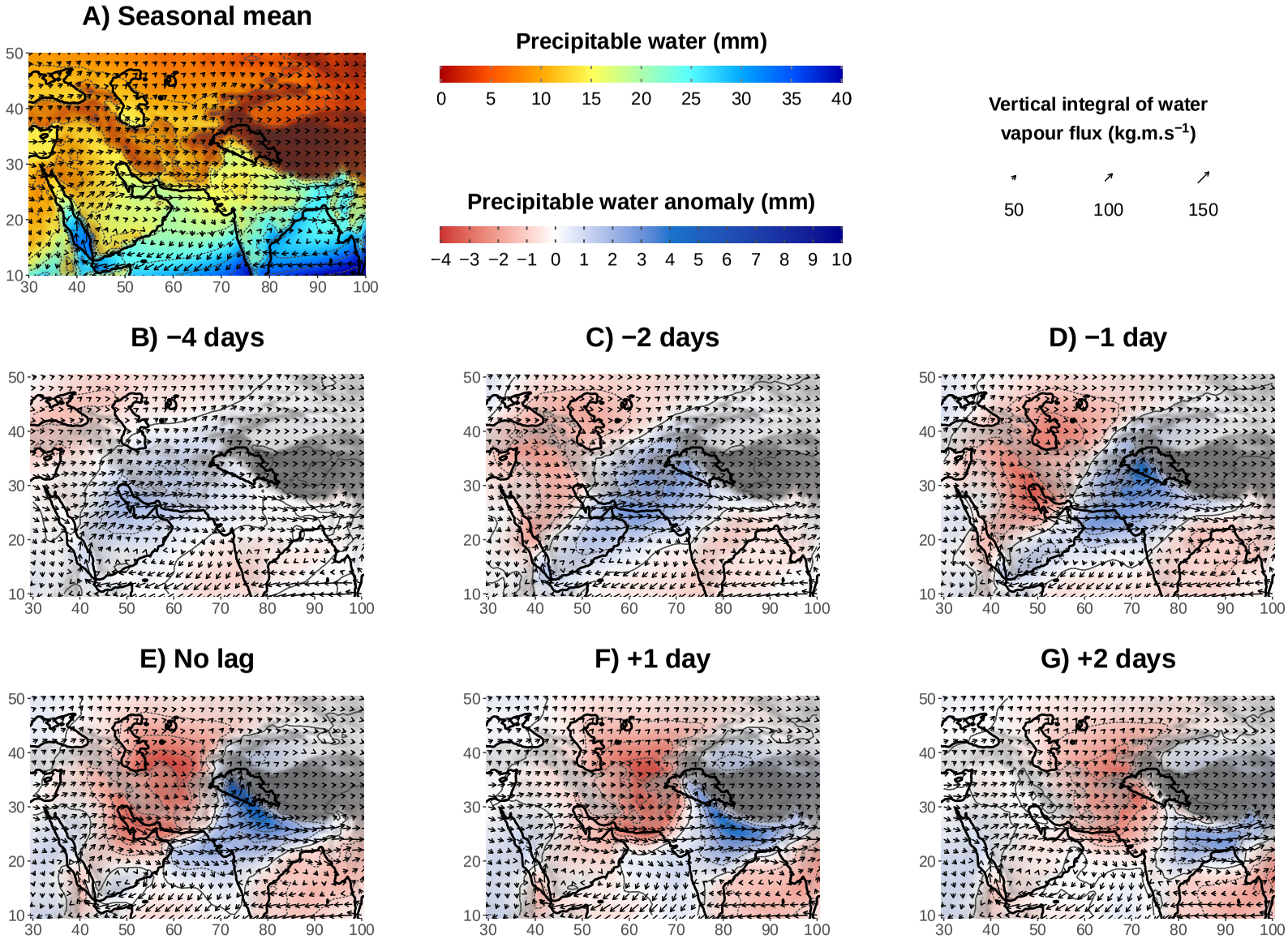

A) Seasonal mean
Precipitable water (mm)
0 5 10 15 20 25 30 35 40
Precipitable water anomaly (mm)
−4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10
Vertical integral of water vapour flux (kg.m.s−1)
50 100 150
B) −4 days
C) −2 days
D) −1 day
E) No lag
F) +1 day
G) +2 days
10 20 30 40 50
30 40 50 60 70 80 90 100

Fig. 4.13 (Previous page) Seasonal mean of precipitable water (colour shading, thin contour lines every 5mm) and vertical integral of water vapour flux (arrows, A). The mean is weighted by the seasonality of occurrence of $W700$ above its 90^{th} percentile. Panels B to G show the lead/lag composite maps, with respect to the 10% highest values of $W700$ (as in Figure 4.2), of precipitable water anomaly (colour shading, thin contour lines every 2 mm) and absolute water vapour flux (arrow). Non-significant anomalies at the level 95% are shown in white.

water in the UIB compared to the mean. The anomaly extends along the Himalayas, towards Northeast India, as well as towards the Arabian Sea. Hunt et al. (2018b) found a similar pattern when investigating extreme precipitation events. This relationship between wind and precipitable water is first explained by the low-level moisture convergence induced by the southerly winds, particularly on the foothills. However, composites of negative lags in Figure 4.13-B, C, and D suggest that anomalously wet air masses advected into the UIB are also responsible for the increase of precipitable water.

Four days before the peak of $W700$, moisture is already building up over the Persian Gulf (Figure 4.13-B). The direction of moisture transport indicates that this moisture is advected from the Red Sea across the Arabian Peninsula. This transport, and how it is enhanced by Mediterranean lows, is described in Chakraborty et al. (2006a) and Mujumdar (2006). This moisture flow passes over the coastal mountain ranges along the Red Sea (Sarawat Mountains, with peak heights ranging from 1000 to 3000 m). Moisture is therefore present at a relatively high altitude, similar to that of the cross-barrier wind in the UIB (700 hPa). The more intense winds at this altitude have the potential to quickly transport the moisture towards the study area.

Two days later (Figure 4.13-D), the positive anomaly of precipitable water has spread from the north of the Arabian Sea to the Indus Plain and the mountains to the west. The south-westerly moisture flux clearly suggests that air masses from the north of the Arabian Sea and the Persian Gulf contribute to the increase in precipitable water, as mentioned by Hunt et al. (2018b). Moisture convergence over the windward side of the mountains also helps to increase moisture at higher altitudes as can be seen over the Suleiman Range, one day before the peak in Figure 4.3-C. Convergence also starts over the UIB, more specifically in the notch formed by the mountain ranges, one day before the peak of $W700$, resulting in a maxima of precipitable water anomaly (Figure 4.13-D).

This analysis shows the existence of a pathway of moisture that sustains moisture content in the UIB. It originates from the Red Sea, crosses the Arabian Peninsula towards the Persian Gulf, continues towards the north of the Arabian Sea, the Indus Plain and ends in the UIB. Previous studies have implied the existence of such a pathway without investigating it further

(Ahmed et al., 2019; Filippi et al., 2014; Hunt et al., 2018b). Part of this circulation, from the Red Sea to the Arabian Sea, is visible in the seasonal mean (Figure 4.13-A) and is driven by the subtropical gyre located on the southeastern tip of the Arabian Peninsula. WDs strengthen this flow and steer it towards the UIB. In some extreme cases, this moisture pathway can form atmospheric rivers (cf. Bao et al., 2006; Zhu and Newell, 1998) that are related to extreme precipitation events along the Himalayas (Thapa et al., 2018). This analysis also suggests that neither the Mediterranean Sea nor the Caspian Sea is important contributors of moisture for the precipitation in the UIB as suggested by other studies (e.g. Dimri et al., 2015).

4.6.3 Evaporation sources

The moisture pathway discussed in the previous Section 4.6.2 does not indicate the sources of moisture, that is, where the water vapour evaporated. That analysis is shown in Figure 4.14, which uses the same composites as Figure 4.13, but for evaporation in the last 24 hours (to avoid the influence of the diurnal cycle) and surface (10 m) winds over the ocean (an important factor in ocean evaporation). The different areas over which the moisture pathway passes are investigated.

The local evaporation in the UIB is quite high compared to the surrounding land, about 1 to 2 mm per day (Figure 4.14-A) and is little affected by the passing of a WD (Figure 4.14-C). While this helps sustaining the generally higher precipitable water in the IUB, it cannot explain the precipitable water anomaly associated with a WD. The rest of the Indus Plain, or the nearby mountains, contribute very little to the moisture, as those areas are arid, and the wetter and higher ground is too cold in winter to generate significant evaporation.

The mean evaporation rate over the Persian Gulf and the north Arabian Sea is about 4 to 5 mm per day (Figure 4.13-A). That is, a day of evaporation represents the quarter of the total precipitable water in the area. This high replacement rate suggests that a large amount of the moisture directly arises from those water bodies, and also emphasizes the importance of the persistence of the moisture pathway to building up moisture. Even stronger evaporation rates are present in the Red Sea, particularly to the north. As mentioned earlier, the importance of the Red Sea comes from the fact that moisture is injected at higher altitude, and therefore more easily transported.

This composite analysis clearly identifies the north of the Arabian Sea, the Persian Gulf, and the Red Sea as the main sources of moisture for the precipitation in the UIB for the first time. However, it does not allow any quantification, nor does it discuss variability. A more complex analysis involving numerical tracking would be needed for this but is beyond the scope of this study.

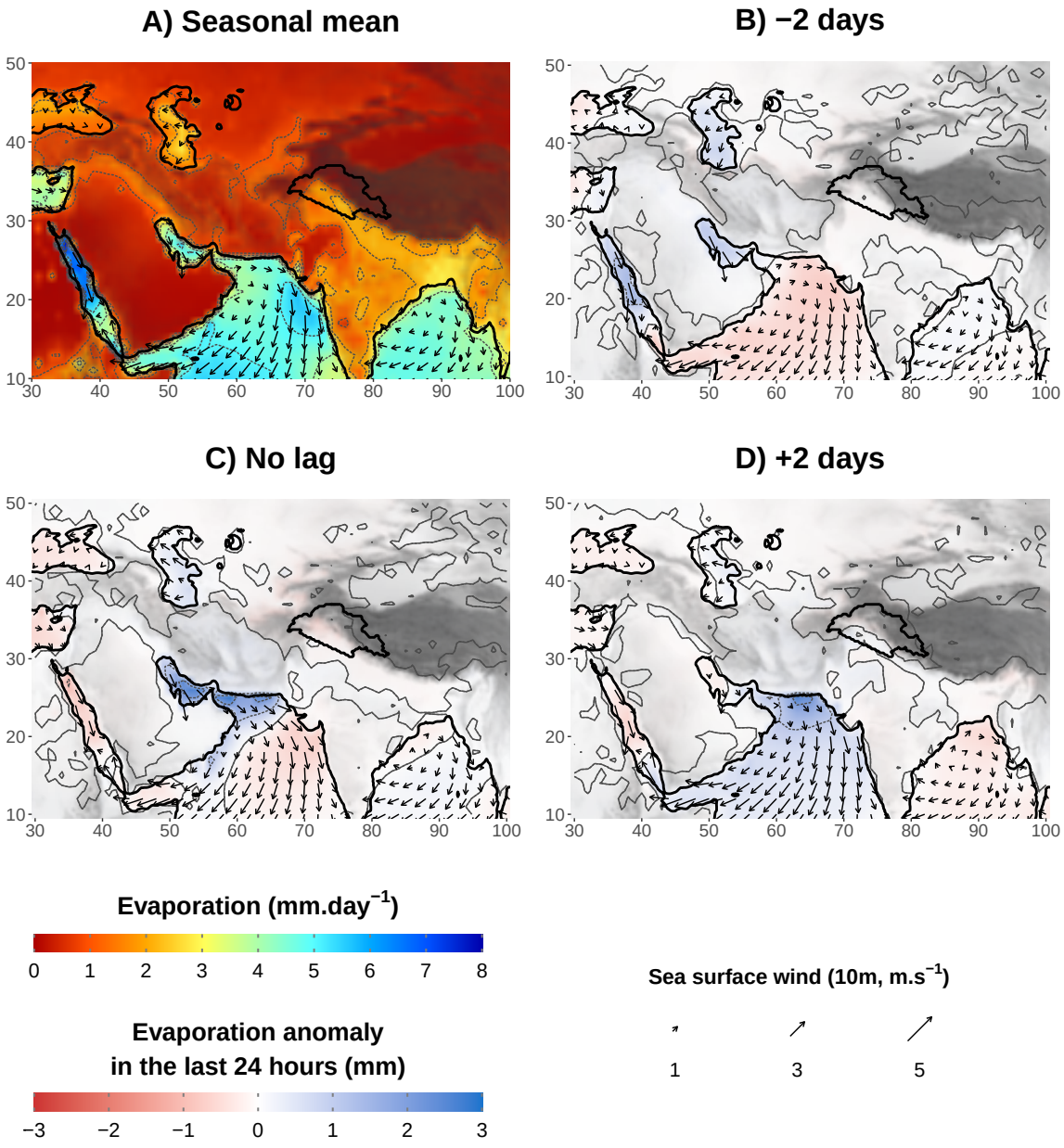


Fig. 4.14 Panel A is similar to Figure 4.13-A but for evaporation (colour shading, thin contour lines every 5 mm) and 10m sea winds (arrows). Panels B to D are similar to Figure 4.13-C, E and G but for cumulated evaporation in the last 24 hours before the peak of $W700$ (colour shading, thin contour lines every 2 mm) and 10m sea winds (arrows).

4.6.4 Impact of WDs on the moisture field

The influence of WDs on the moisture pathway to the UIB has already been discussed (Section 4.6.2), but other impacts are evident from Figure 4.13 and 4.14. Two days before the peak of $W700$, a negative anomaly of precipitable water propagates from Mesopotamia into the Persian Gulf (Figure 4.13-C), and reduces the moisture supply from the Red Sea. This circulation is the effect of a surge of cold, dry air at the rear of a WD. The negative anomaly eventually propagates to the north of the Arabian Sea after the peak of $W700$ (Figure 4.13-E, F, and G), but tends to weaken. This weakening can be explained by the large evaporation anomaly present over the seas as the dry air progresses (Figure 4.14-B, C, and D). Evaporation rates almost double compared to the seasonal mean (Figure 4.14-A), which is enough to eliminate the precipitable water anomaly in a day. This evaporation anomaly can be explained by both the increase of surface wind speed induced by the WD (arrows on Figure 4.14) and the very low surface dew point of the air blown from the land but is not further investigated.

A second but connected area of negative anomaly of precipitable water is present over Central Asia and the Caspian Sea (Figure 4.13-C and D). It also relates to cold and dry air being advected by the WD from northern latitudes. When $W700$ peaks, that cold and dry air mass invades the mountains to the west of the Indus Plain (Figure 4.13-E, see also Figure 4.3-B and D). Unlike the other air mass passing over the seas, the encounter with the relief further increase the moisture anomaly through a Foehn effect. This effect is strongest after the peak of $W700$, as the mean moisture flux over the Indus Plain veers eastward. The continental dry air descends from the Suleiman Range and the Hindu Kush, replacing the maritime moist air (Figure 4.13-F and G). As the UIB is cut from its moisture supply, the precipitable water anomaly becomes negative (Figure 4.13-G), which severely limits moisture transport and precipitation despite the remaining presence of cross-barrier wind. Similarly, Figure 4.6 indicates that the passing of a WD noticeably reduces moisture contribution ($Q700$) several days after the peak of $W700$.

Hence, a WD has two opposite effects on moisture content in the UIB: moistening before its passing, and drying after. Therefore, another ensuing WD may be inhibited, in terms of precipitation, by the dryer conditions dominating in the Indus Plain and the UIB, suggesting a negative feedback effect. However, one could also imagine a positive feedback, where a quick succession of WDs provides a continuous moisture supply from the maritime sources by inhibiting the dry air invasion. The interaction between WDs is not further investigated here and is likely subject to high variability and dependant on larger and smaller-scale atmospheric circulation.

After the peak of $W700$, the positive anomaly of precipitable water moves along the Himalayas towards Northeast India where it enhances cross-barrier moisture transport and orographic precipitation in another notch that is formed by the relief, enhancing orographic precipitation there (not shown). WDs are not as effective in this area, compared to the UIB, first because the upper-level disturbances have mostly disappeared due to the interaction with the relief, and because convection is a more important driver of precipitation (Mahanta et al., 2013; Mannan et al., 2017; Tinmaker and Ali, 2012).

Finally, the north-easterly trade winds that blow over most of the Arabian Sea are also affected by WDs: they are weaker ahead of the WD, leading to a decreased evaporation (Figure 4.14-B), which also slightly decreases precipitable water along 10°N (Figure 4.13-B to E). On the contrary, trade winds intensify after the passing of the WD with an increased evaporation (Figure 4.14-D) associated with a slight increase in precipitable water to the south (Figure 4.13-F and G).

4.6.5 WD variability regarding moisture

This section discusses how various characteristics of WDs influence moisture content in the UIB ($Q700$). The most obvious is the intensity of the low-level cross-barrier winds, defined by $W700$ (cf. concomitant peak of $W700$ and $Q700$ in Figure 4.6): the correlation between $W700$ and the seasonal anomaly of $Q700$ reaches 0.45 when considering all winter time steps. Cross-barier winds increase moisture content at 700 hPa by lifting moisture from lower levels, but also because it is generally accompanied by south-westerly winds in the Indus Plain supplying the UIB with moisture from the nearby seas (cf. Figure 4.13). Since $W700$ is related to the upper-level intensity and the thermal structure of a WD, both also impact $Q700$. However, the shape and direction of a WD has an even greater impact by acting on the balance between the supply of maritime moist air and the intrusion of continental dry air as discussed earlier (Section 4.6.4). Two sub-selections (25% each) of the time steps with $W700$ above the 90^{th} percentile are created. These selections are different than the one used for the analysis of the thermal structure (Section 4.4.3, Table 4.1), but also use a quantile regression. One of the selections characterises low $Q700$, and the other one high $Q700$. The two predictors of the quantile regressions are the months, to remove the impact of seasonality, and the difference of geopotential anomaly between the grid points 60°E-36°N and 70°E-36°N. This latter predictor fixes the longitudinal gradient of geopotential across the UIB, and in that way the position of the mean WD, to avoid over-representing timesteps before (after) the maximum of $W700$ in the group of high (low) $Q700$ (cf. Figure 4.6). Various composites are derived from theses selections (Table 4.2, Figures 4.15, 4.16, and 4.17).

Table 4.2 Same as Table 4.1 but using the selections of low $Q700$ and high $Q700$ defined in Section 4.6.5

	Precipitation ($mm \cdot 3hrs^{-1}$)	$W700$ (ms^{-1})	$Q700$ ($kgkg^{-1}$)	$\widehat{W700}$ (ms^{-1})	$Z300$ (ms^{-1})	dZa_UP (ms^{-1})	dZa_LW (ms^{-1})
Lower $W700$	5.9	3.8	2.3×10^{-3}	3.6	1.6	1.1	0.2
Higher $W700$	14.8	4.4	4.2×10^{-3}	3.8	1.1	1.7	0.4

Table 4.2 gives the mean values of different variables investigated for the two selections. The discrimination is successful, as $Q700$ is almost twice as high in the selection characterising higher $Q700$. A difference in $W700$ is evident, but is too small to explain the difference in $Q700$, indicating that cross-barrier winds are not the main driver of moisture variability. Interestingly, the difference in $\widehat{W700}$ is smaller, indicating that non geostrophic processes are mostly responsible for the difference in $W700$. Increased convective activity due to moister air masses can be one of these processes (Dimri et al., 2015). Small differences are also evident in the WD structure. The geostrophic wind at altitude ($Z300$) is stronger in the case of lower $Q700$, but the stronger thermal gradient below it tends to reduce its impact at lower altitude. As suggested in Section 4.4.3, the increased warming induced by latent heat release in the case of higher $Q700$ prevents the cold air accompanying the WD from entering the UIB, which enhances low-level winds ($W700$). As expected, precipitation rates are much higher in case of higher $Q700$. These differences are investigated further using composite maps.

Figure 4.15 shows precipitable water anomaly and absolute vertical integral of water vapour flux as in Figure 4.13, but for the two selections. As expected, the two selections exhibit very different moisture conditions when $W700$ is maximum (Figure 4.15-C and F). In case of high $Q700$, a high positive anomaly of precipitable water is present from the Arabian Sea to Northeast India, with a maximum over the UIB, indicating a sustained supply in moisture. By contrast, a negative anomaly of precipitable water is present over the Indus Plain in the case of low $Q700$, showing that continental dry air is already invading, cutting any potential moisture supply from the Arabian Sea and the moisture pathway. A small positive anomaly remains in the UIB, due to cross-barrier winds. Interestingly, no evident difference in moisture flux direction is visible between the two panels (Figure 4.15-C and F).

The explanation of the differences in moisture content lies in the circulation history. In case of high $Q700$, moisture over the Persian Gulf builds more importantly than in the average case four days before the peak of $W700$ (Figures 4.13-B and 4.15-D), due to stronger

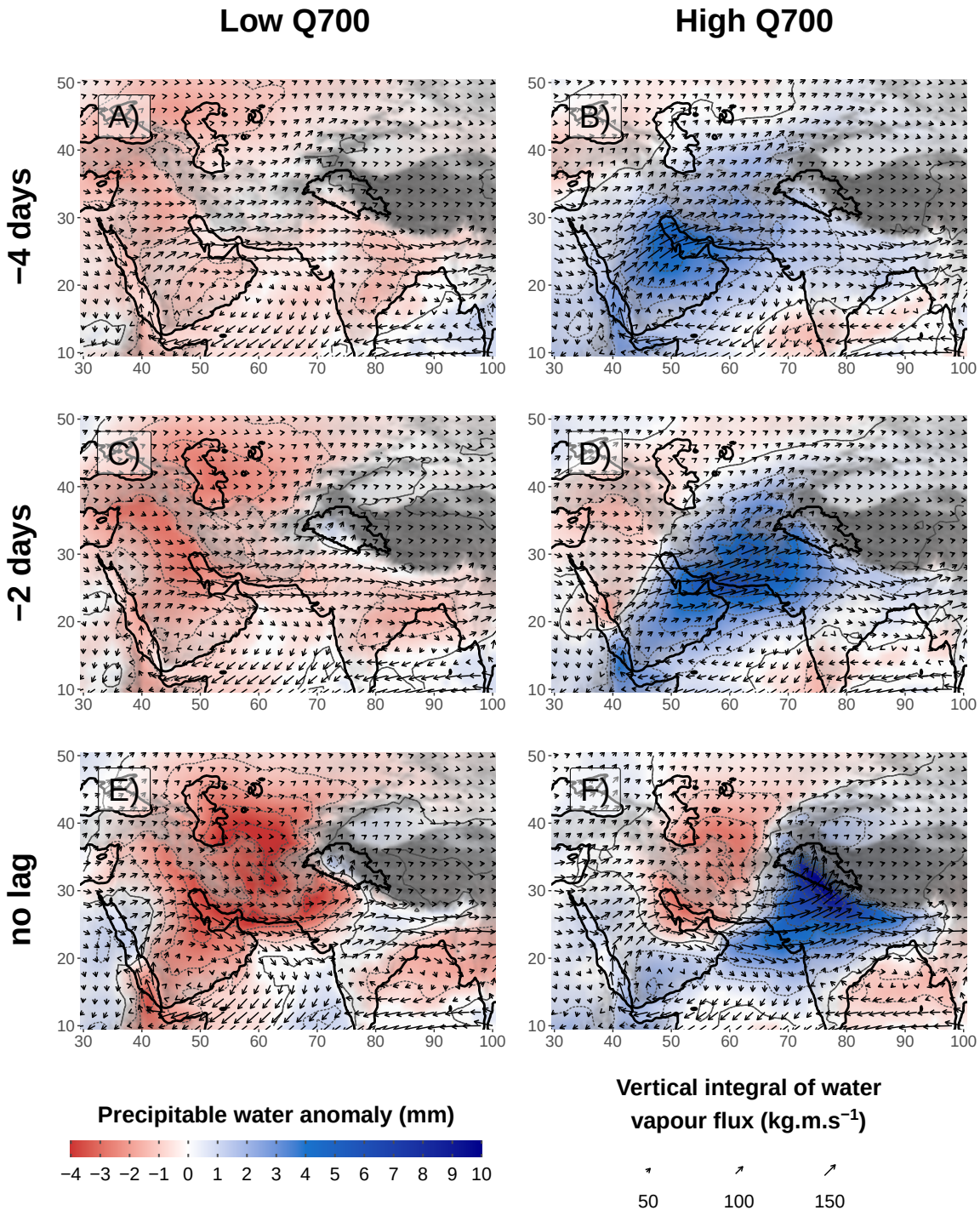


Fig. 4.15 Same as Figure 4.13 (B, C and E), but the composites are based on the lowest (highest) values of $Q700$ in panels A to C (D to F). See the selection definition in text.

moisture transport across the Arabian Peninsula. This moisture anomaly moves towards the Indus Plain, the mountain ranges to the west, and even towards parts of Central Asia two days later (Figure 4.15-E), pushed by a strong southerly component of the moisture flux. The larger extent of the precipitable water anomaly than in the average case (Figure 4.13-C) effectively delays the intrusion of continental dry air (Figures 4.15-F and 4.13-C). In the low $Q700$ case, no such positive anomaly is present (Figure 4.15-A and B): the WD is not accompanied by a strengthening of the moisture transport along the pathway, and the southerly component of the moisture flux is almost absent. Instead, negative anomalies of precipitable water build up over Mesopotamia and Central Asia (Figure 4.15-B) and move eastward and south-eastward respectively towards the Indus Plain (Figure 4.15-C). These patterns show that the balance between moisture advection ahead of a WD and dry air intrusion at the rear may vary significantly. The pattern does not change if $W700$ is fixed in the quantile regression, suggesting that this mode of variability is more important than the intensity of $W700$ in explaining $Q700$ variability (at least as long as $W700$ is intense enough, due to the definition of the composites).

Finally, trade winds in the Arabian Sea also vary between the two selections: they are stronger (weaker) in case of lower (higher) $Q700$ (arrows, Figure 4.15). Figure 4.13 suggests that trade winds are impacted by the passing WD, but trade winds can also impact the moisture field outside of WD influence. Stronger trade winds transport the moisture evaporating from the Arabian Sea to the Southern Hemisphere, while weaker trade winds allow for a build-up of moisture.

The mean WD from each selection has different shapes and motion, as defined by the geopotential height at 500 hPa (Figure 4.16). At the peak of $W700$, the mean WD is located at the same longitude, indicating that the predictors of the quantile regression are used successfully (Figure 4.16-C and F). Interestingly, the latitude of the WD in case of high moisture is lower than that in the dryer case. More surprisingly is the deeper geopotential anomaly in case of low $Q700$, despite a slightly lower $W700$ (cf. Table 4.2). The difference further amplifies at 300 hPa (not shown). The thermal structure only partly compensates for the difference of geopotential gradient at higher altitude as discussed for Table 4.2.

The difference in moisture flux in the days before can be explained by the lagged composites. In the case of high $Q700$, a weak but well defined precursor is present over the Levant four days earlier and quickly moves eastward toward the UIB (Figure 4.16-B and D). This precursor is similar to the one found in the composite map for high $W700$, which showed a stronger influence from the thermal structure (Figure 4.12-B). However, differently to this map, a much weaker positive anomaly of geopotential height is present over Eastern Europe, without evidence of a Rossby wave breaking (cf. the purple contour line). The low

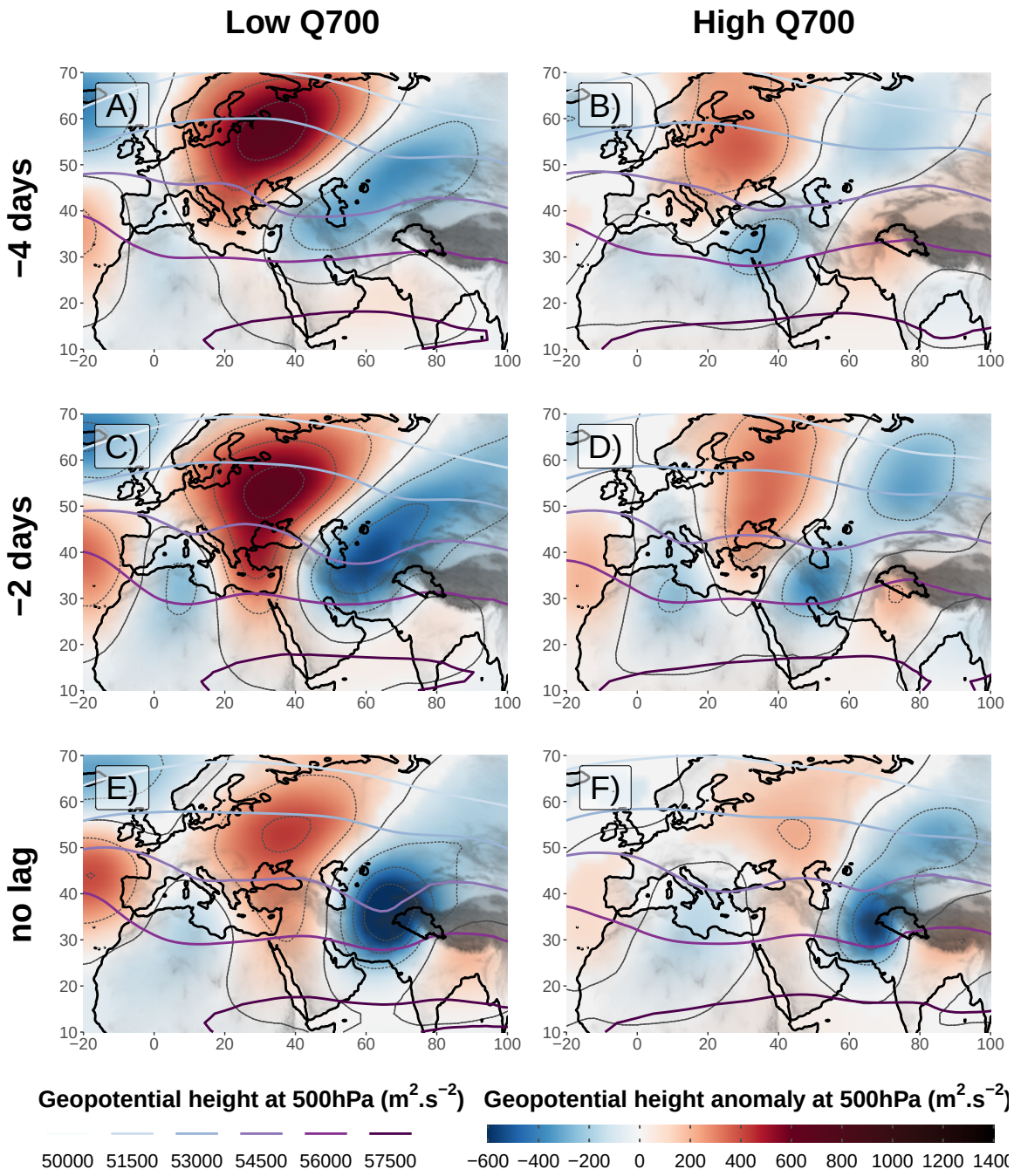


Fig. 4.16 Same as Figure 4.10 (A and B), but for composites based on the same selection as in Figure 4.15.

latitude position of the WD within the jet and the zonal gradient ahead of it explains the stronger interaction with the moisture pathway and the better moisture supply to the UIB. Possibly, the higher latent heat release over the mountains to the west of the UIB, due to the higher moisture content, further enhances the low-level wind circulation and contributes to the development of the WD.

In case of low $Q700$, the WD develops from the south-western tip of a positively tilted trough positioned over Central Asia four days before the event (Figure 4.16-A). A strong gradient of geopotential is present to the north-west of the trough. As suggested for the selection of lower $W700$, this gradient enables an equatorward motion of continental cold and dry air, which allows for the development of a baroclinic wave. However, since the trough is too far north, it does not interact with the moisture pathway and instead promotes the surge of dry continental air towards the Indus Plain.

Finally, the two types of WD investigated here also have a different jet signature (Figure 4.17). In the low $Q700$ case, four days before the peak of $W700$, the Arabian Jet streak is anomalously strong and already extends to the UIB (Figure 4.17-A). The developing WD progressively pinches the jet, but remains in the left entrance of the jet streak as it progresses eastward (Figure 4.17-B). This area is characterised by altitude convergence and downward motion, which could explain the development of the upper-level low, through stretching of the vorticity, but also the drying of the atmosphere. It also inhibits the uplift in the UIB. The distinct outflow signature in wind speed anomaly to the northeast of the WD (cf. Figure 4.9-C) is mostly absent. Instead, the jet is funnelled right over the Tibetan Plateau. Consequently, the habitual weakening of the East Asian Jet is not present, and instead intensifies several days after the peak of $W700$ (Figure 4.17-C).

In the case of high $Q700$, the jet exhibits opposite features (Figure 4.17-D to F), more similar to those of the average case (Figure 4.9). The positive anomaly of the Arabian Jet streak is moderate, but the WD is located on the left exit of the jet as it develops, which is favourable to the uplift in the UIB and to the low-level cyclonic circulation. The outflow (i.e. the positive wind speed anomaly to the north-east of the UIB) is also more intense, at the expense of the core of the jet to the south (Figure 4.17-E). Consequently, the East Asian Jet is severely weakened (Figure 4.17-F).

In conclusion, it is clear that the North Arabian Sea, the Persian Gulf, and the Red Sea are the main contributors of moisture in the UIB. These three source regions are linked by a moisture pathway, which allows the build-up of moisture. The passing of a WD enhances this moisture pathway and steers it toward the UIB. Intrusion of dry continental air at the rear of a WD disrupts the moisture supply to the UIB. Large variability exists in the ability of WDs to steer moisture to the UIB, depending on the development history of the WDs and

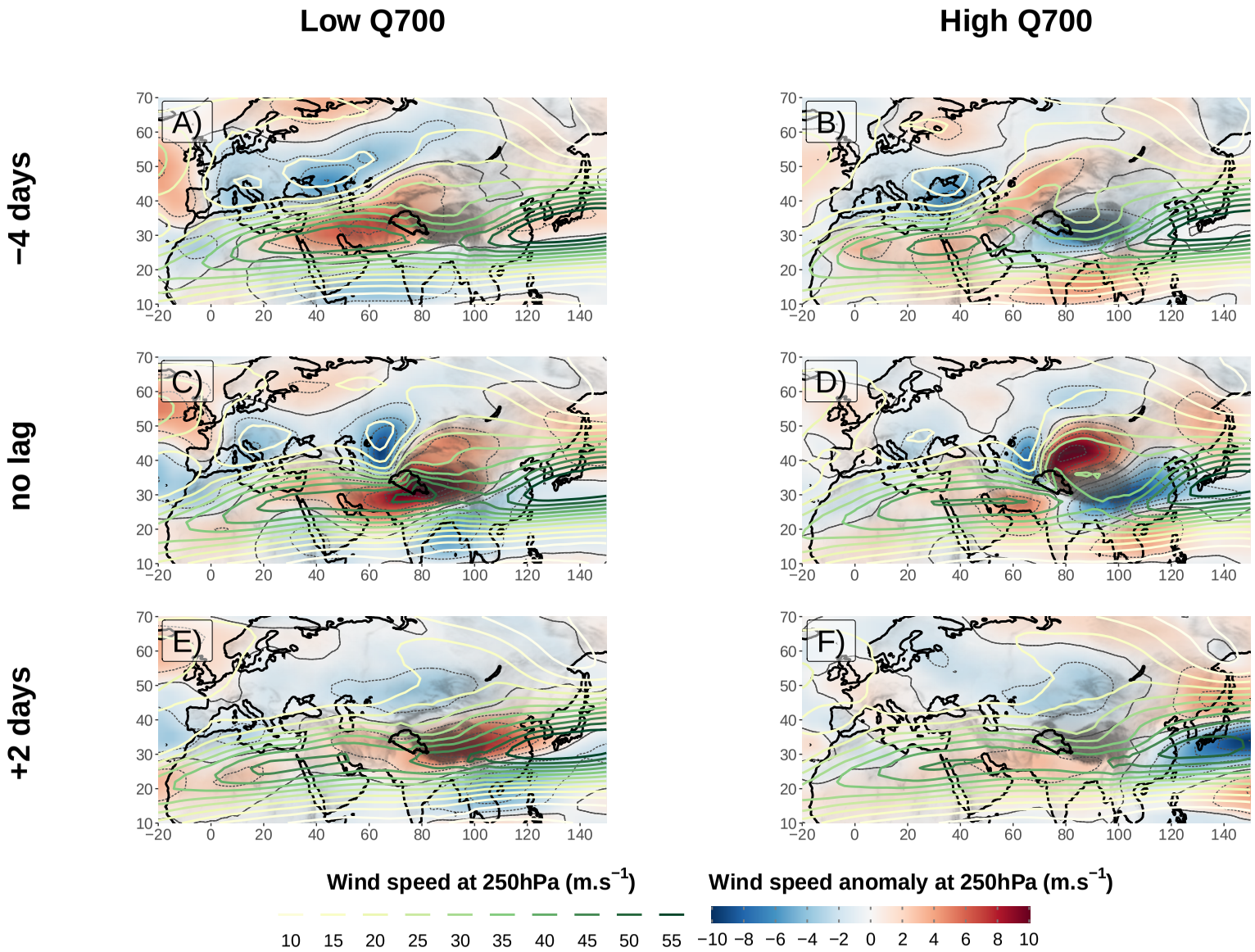


Fig. 4.17 Same as Figure 4.9 (A, C and D), but for composites based on the same selection as in Figure 4.15.

the Arabian Jet position. Particularly, the anticyclonic anomaly over Eastern Europe has a double role: while it provides a corridor for cold air that fuels the development of WDs, this cold and dry air advection can also inhibit precipitation formation.

4.7 Understanding precipitation seasonality

4.7.1 Active season

Precipitation reaches a maximum in March after a very dry autumn (October-December) and a less dry minimum in May (Chapter 2). However, the drivers of such seasonality are not fully known, while they would help to understand the climate model biases discussed in Palazzi et al. (2015). For example, the seasonality of WDs established by tracking suggests a peak in January with an equally lower activity in autumn and spring (Hunt et al., 2018a). The distinction of precipitation variability into wind and moisture contributions and the analysis of their respective synoptic variability developed here help to understand the seasonality of each variable.

Moreover, an even simpler question is still unanswered: in which season is precipitation mostly driven by WDs? Hunt et al. (2018a) argue that WDs occur in all seasons, although no strong WDs develop in summer, and the weaker summer WDs are significantly less frequent than in winter. They also suggest that the WD active season lasts from December to April (Hunt et al., 2018a). Dimri et al. (2015) considered the same period for their analysis, while Dimri and Chevuturi (2016) mention the importance of WDs during the pre-monsoon and monsoon period. Here, the period October to May is considered, which is much longer than what has been considered previously in the literature, which needs to be justified. To address this question, a composite is calculated including 10% of the 3-hourly timesteps with the highest precipitation rate for each month. The precipitation threshold is therefore dependant on the month. Note the difference with the previous composites that used $W700$ for the selection. Figure 4.18 depicts the vertical structure of the atmosphere as in Figure 4.3-A and gives the fraction of total precipitation for each monthly composites.

For all months between October and May, precipitation events corresponds to the presence of a mean negative anomaly[4] of geopotential height at high altitude (300 hPa) around 66°E, which characterises an extra-tropical disturbance, and is used as a basis for the definition of "winter" in this study. By contrast, during June and September a much weaker negative geopotential anomaly is present at lower-level (600 hPa), while in July and August, no

[4]The anomalies are computed by removing the first four harmonics of the annual cycle, and thus not relative to the monthly mean, cf. Section 4.4.1

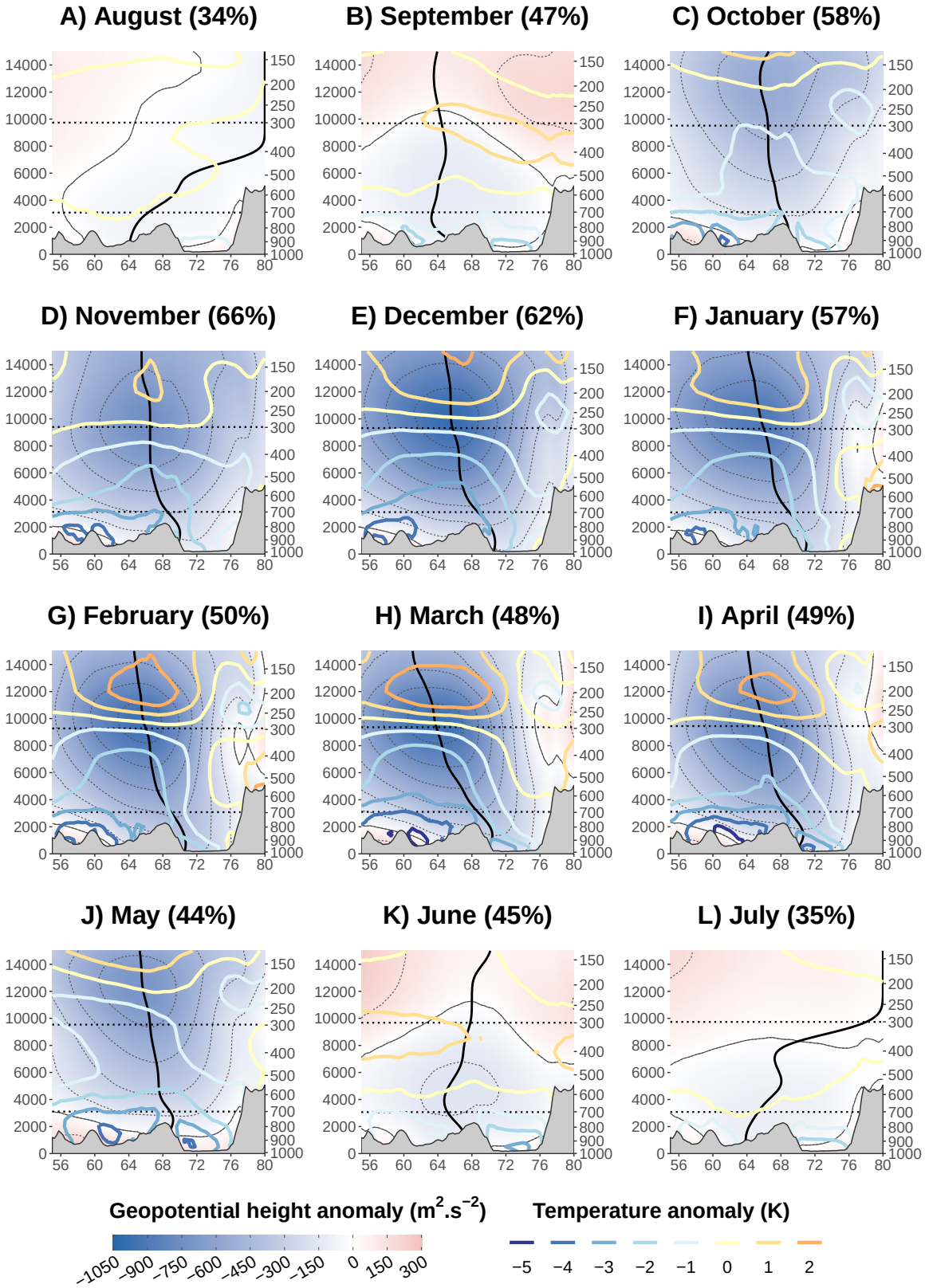


Fig. 4.18 Composite cross-section as in Figure 4.3-B. The composites are based on the 10% highest 3-hourly precipitation rates for each month. The percentage of monthly precipitation accounted for by the selections is indicated in parenthesis.

substantial anomalies are evident. Yet, some differences are notable between the winter months. For example, the anomaly is located at higher altitude, around 200 hPa, in October and May (Figure 4.18-C and J). This higher altitude can be explained by a higher mean tropopause during those months. A lag analysis for those two months (not shown) suggests that the geopotential anomaly generally comes from the north or north-west, while almost no tilt is evident, indicating a weaker baroclinicity. These disturbances may be more similar to a cut-off low (Dimri and Chevuturi, 2016), than to a traditional fast moving WD, although of the same extra-tropical origin, and thus unrelated to the monsoon processes. For all other winter months (November to April), the similarities to each other and to Figure 4.3-B are striking, except for a stronger tilting in spring, particularly in March.

4.7.2 Results from the regressions of precipitation

Figure 4.19 presents the seasonality, centred in winter, of the different variables discussed here. Chapter 3 showed that the prediction of precipitation by moisture transport (the product of wind and specific humidity) at 700 hPa and 850 hPa represents the seasonality of precipitation well. The winter seasonality is even better reproduced here, with small changes to the regression technique (e.g. the regression is limited to the period October to May and a smaller area, all directions of moisture transport are used, as well as an unbiased optimisation for positive values, cf. Section 4.3) and the use of moisture transport at 700 hPa as the only predictor (the two black lines, Figure 4.19-A). The bias only grows at the edge of the selection of winter months (early October and end of May) and in summer (which is not used in the regression), when the contribution of the moisture transport at 850 hPa becomes more important (cf. Chapter 3). The seasonality of wind and moisture contributions ($W700$ and $Q700$, respectively red and blue dashed lines in Figure 4.19-A) is also very similar to that found in Chapter 3. Note that negative values of $W700$ are included in the seasonality although they do not contribute to precipitation; doing so has a limited impact on the seasonality of precipitation. $W700$ is maximal in February; the peak is relatively symmetrical, with lowest values in October and June. $Q700$ drops quickly in October to reach a minimum between December and January. Thereafter, it rises steadily. While $W700$ is the driver of the winter peak of precipitation, $Q700$ modulates its effects, particularly delaying the precipitation peak by a month and sustaining precipitation in October and spring. The effect of moisture availability thus partly reconciles precipitation seasonality with that of WDs.

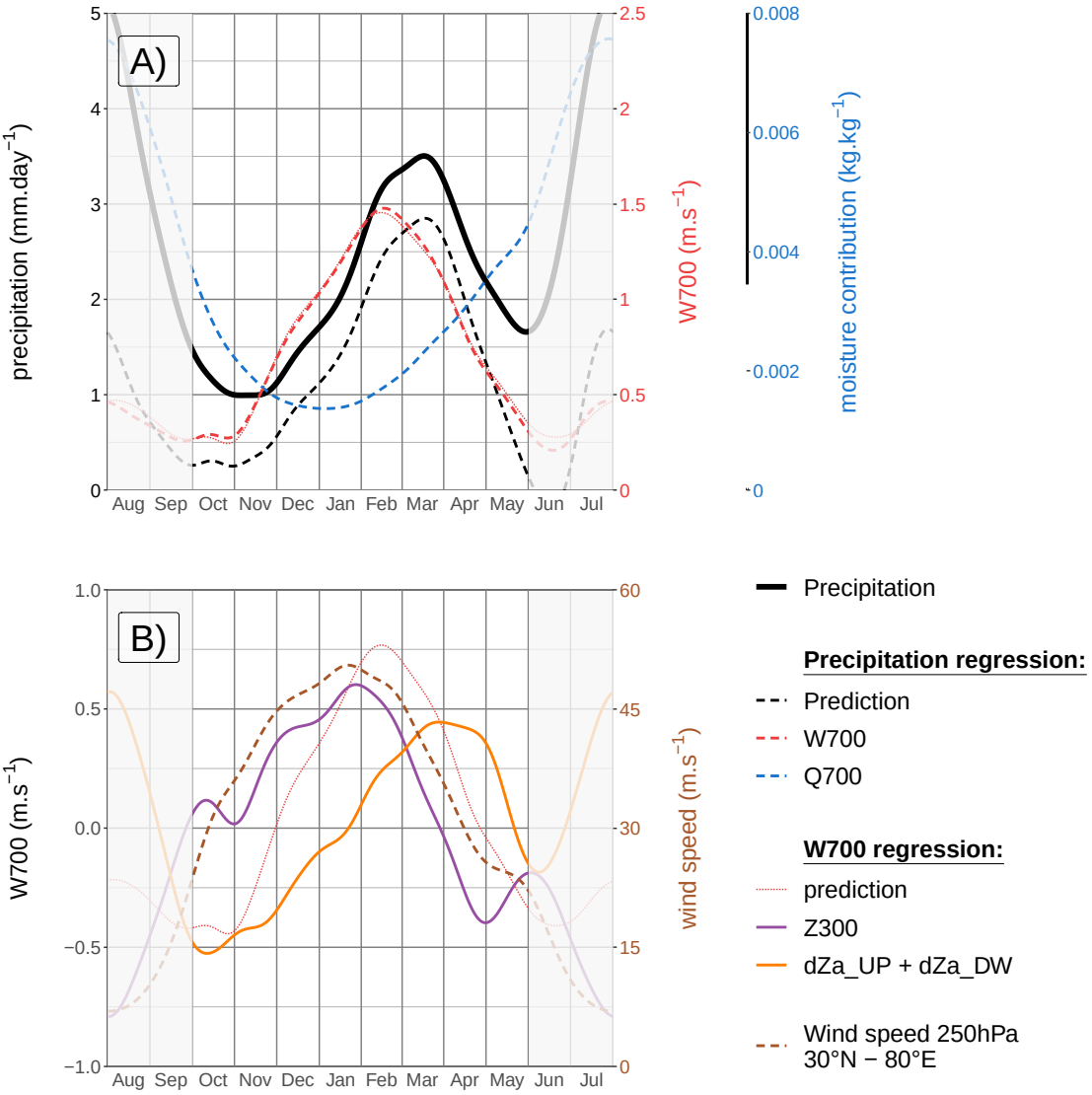


Fig. 4.19 In panel A, seasonality of precipitation, the prediction and the contributions to precipitation based on the regression with moisture transport (cf. Figure 4.1). In panel B, seasonality of the contributions to $W700$ based on the regression with geopotential heights and thicknesses (cf. Figure 4.5). The prediction of $W700$ ($\widehat{W700}$, fine dotted red line) is present in both panels, but without the offset in panel B, so it is the exact sum of the two contributions. The seasonality of the wind speed at 250 hPa, 30°N and 80°E is also represented in panel B.

4.7.3 Wind contribution seasonality

The seasonality of $W700$ is further investigated using the result of its regression with geopotential height and thickness (cf. Section 4.4.2). The seasonality of the prediction fits almost perfectly that of $W700$ (red lines in Figure 4.19-A), allowing the analysis of the seasonality of the different contributions (Figure 4.19-B). The contribution of geopotential height at 300 hPa ($Z300$, purple line) shows a peak at the end of January, about a month earlier than the wind. $Z300$ is related to geostrophic wind at 300 hPa and particularly its south-westerly direction over the UIB (cf. gradient in Figure 4.5-B). Hence, it is not particularly surprising that $Z300$ seasonality follows that of the SWJ wind speed as shown by the dashed brown line (Figure 4.19-A). The seasonality of the SWJ intensity is also in agreement with the seasonality of WD found by Hunt et al. (2018a), particularly with a peak intensity in January and a decrease in spring. This relation between the mean SWJ intensity and the synoptic circulation at 300 hPa highlights the importance of the SWJ intensity as a driver for WDs. A stronger SWJ also implies stronger windshear and a stronger ageostrophic circulation (cf. Sections 4.4.3 and 4.5), which allows for the development of potentially deeper lows.

The SWJ seasonality is further described in Figure 4.20. Panel A indicates the mean position and intensity of the SWJ at 70°E, where the wind speed anomaly is the strongest when a WD passes (cf. Figure 4.9-C). The seasonality exhibits two different behaviours. When WDs are active, the SWJ has a stable mean position between 28 and 30°N. It undergoes an intensification and peaks at the end of January before weakening. Starting from May, the SWJ strength stabilises while moving northward as the monsoon circulation is established. It reaches its northernmost position in July-August before coming back to its winter position. While the results are mostly in agreement with Schiemann et al. (2009), the clear decoupling between intensity and latitude contradicts one of their hypothesis, i.e. that the jet intensity is driven by the change in latitude[5]. Instead, Figure 4.20-A suggests that the thermal gradient is the primary driver. In the context of WDs, both position and strength of the jet are important: the position enables the propagation of WD towards the UIB, while the jet speed determines WDs potential intensity. Note that the wiggling of the seasonality of $Z300$ in October and May (Figure 4.19-B) corresponds to the moment the SWJ reaches the latitude of the UIB, causing a small intensification of the upper-level wind, but not of the WD activity as the jet is too far north.

The SWJ at 70°E exhibits a particularly symmetrical seasonality, but behaves slightly differently at other longitudes. Figure 4.20-B shows that the eastern part of the SWJ, the East Asian Jet, is stronger during the early part of the winter season, and extends south of

[5]This hypothesis is based on the thermal wind equation, where a change in the Coriolis parameter changes the wind speed, even if temperature gradients are identical

the Tibetan Plateau. As a result, upper-level wind speed increases in the UIB as indicated by the higher value of $Z300$ in December and January in Figure 4.19-B. However, in this configuration, the UIB is the left entrance of the East Asian Jet, which inhibits uplift. In fact, when precipitation occurs during this period (see the $40\,\mathrm{m\,s^{-1}}$ green contour line in Figure 4.20-B), the SWJ has a similar shape as in the case of a WD's thermal structure inducing lower $W700$ (Figure 4.11-A). Hence, the contribution of the thermal structure (dZa_UP and dZa_LW) is negative during that part of the season (orange curve, Fig 4.19-B).

By contrast, during the later part of the winter season, the East Asian Jet weakens, while the Arabian Jet separates from it (Figure 4.20-C). When a WD passes, the Arabian Jet extends toward the UIB (green contour line), and the study area becomes in the left exit of the jet, as in the case of higher $W700$ (Figure 4.11-B). Consequently, the contribution of the thermal structure is positive during that part of the season (orange curve, Fig 4.19-B), and peaks in March-April. Furthermore, the increase in moisture content and the associated latent heat

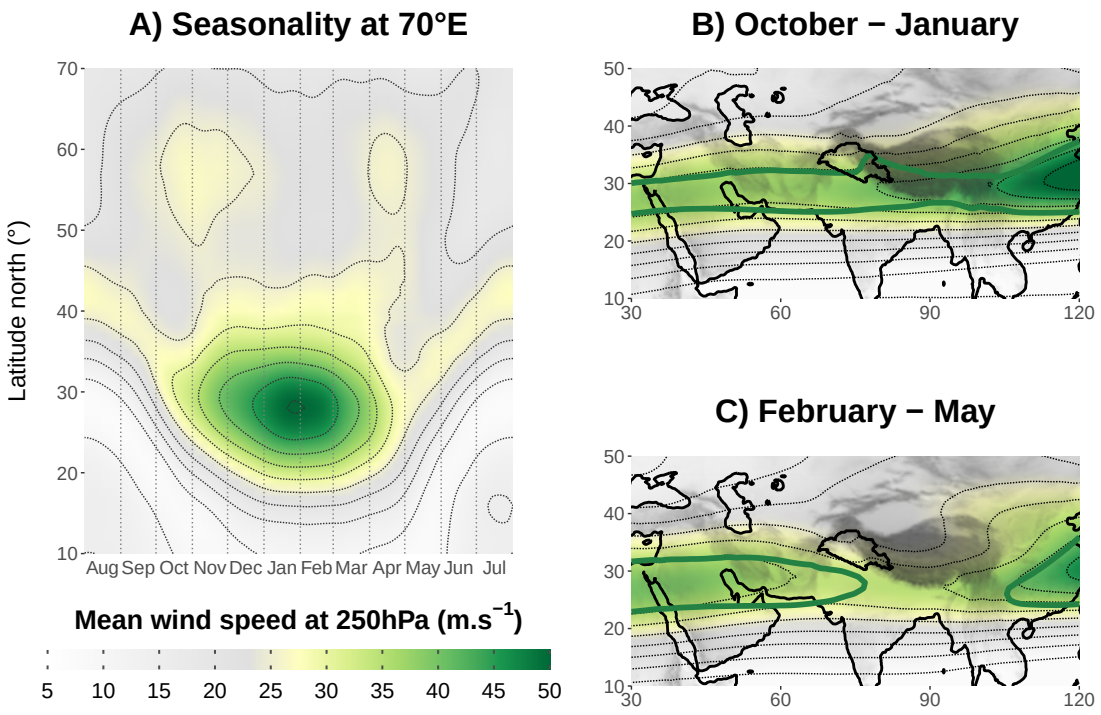


Fig. 4.20 Hovmuller representation of the seasonality of the wind speed at 250hPa along 70°E (thin contour lines every $5\,\mathrm{m\,s^{-1}}$, A). In panel B (C), map of the mean wind speed at 250 hPa during the earlier (later) part of the winter season. The green contour line in panel B (C) represents the $40\,\mathrm{m\,s^{-1}}$ contour line in case precipitation is above its 90^{th} percentile for the earlier (later) part of the winter season.

release during these months likely impacts the thermal structure contribution, but remain to be quantified positively.

The difference of seasonality of the Arabian Jet and the East Asian Jet is likely related to the effect of surface heat sources and sinks, particularly that of the Tibetan Plateau. Due to the seasonality of solar radiation, the Tibetan Plateau acts as a heat sink in autumn and a source in winter and drives the intensity of the East Asian Jet (Schiemann et al., 2009; Yanai et al., 1992). Similar processes lead to the variation of the Arabian Jet, but the lower altitude of the heat sources and sinks lead to a further delayed impact on the Arabian Jet. Hence, the East Asian Jet peak intensity occurs closer to the solar radiation minimum than that of the Arabian Jet.

4.7.4 Moisture contribution seasonality

The seasonality of $Q700$ (Figure 4.19-B) could be explained by various processes, which can be investigated:

- The temperature and the maximum quantity of moisture the air mass can hold through the Clausius-Clapeyron relationship as suggested by Cannon et al. (2016b).
- The ability of WDs to steer moisture towards the UIB.
- The strength of the moisture pathway from the Red Sea to the Arabian Sea and the build up of moisture it allows.
- The evaporation rate in the different moisture sources.

Evaporation over the ocean surfaces is not responsible: the highest rates during winter occur in December-January and reach a minimum in October and May (not shown), which is opposite to the seasonality of $Q700$. In fact, the higher evaporation in December-January is in part explained by the drier conditions, while higher wind speed, mainly the trade winds in the Arabian Sea explains another part. In contrast, the local evaporation rate in the UIB (evapotranspiration) better corresponds to $Q700$ seasonality: it drops in autumn to reach a minimum in January and then rises until April. Precipitation through soil moisture and surface temperature explains that seasonality. However, the amounts involved are not high enough to explain $Q700$ seasonality.

Sea surface temperature is another important driver of evaporation and impacts surface moisture content (Chaboureau et al., 1998). However, it reaches a minimum in February for the Red Sea, the Persian Gulf, and the Arabian Sea, which are considered to be the sources of the UIB moisture, and has a two months delay compared to the $Q700$ minimum. Surface

temperature in the UIB exhibits a seasonality closer to that of $Q700$, with a relatively sharp minimum in January. However, this seasonality does not explain the higher moisture content in spring than in autumn, as well as the origin of the excess of moisture.

Finally, the mean moisture transport and the impact of WDs are investigated in Figure 4.21. This figure combines information about the intensity of the monthly mean moisture flux (green contours), the moisture flux direction (arrows) one day before a precipitation event (defined as precipitation above the 90^{th} percentile for each month, as in Figure 4.18), and the resulting anomaly of water vapour flux intensity (colour shading).

The monthly mean highlights the two important features of the wintertime circulation already discussed: the trade winds, and the westerly moisture pathway from the Red Sea to the Arabian Sea. The trade winds blow from the Arabian Sea and the Bay of Bengal to the Indian Ocean across the equator. They quickly established in October to peak end of December. Then, they weaken from January to April, before being replaced by the opposite summer monsoon circulation in May. This seasonality strikingly matches $Q700$ seasonality. Indeed, stronger trade winds imply stronger dry advection from the continent, which depletes moisture content in the whole area, while the moisture evaporating from the Arabian Sea is transported southward. In contrast, weaker trade winds allow the build-up of moisture.

Meanwhile, the moisture pathway from the Red Sea establishes more progressively than the trade winds, following the southward migration of the subtropical anticyclonic gyre (Arabian high). The part across the Arabian Peninsula establishes first in November, before extending eastward throughout the whole winter seasons. In May, as the subtropical high moves north again, moisture supply from the Red Sea disappears.

While both the trade winds and the westerly moisture pathway are characteristic of the wintertime circulation, there seems to be a competition between the two, so that the westerly transport can only develop once the trade winds weaken. A similar relationship is described at the synoptic scale during the passing of a WD (cf. Figure 4.13). In fact, the change in balance between the two circulations may be related to a change in WDs characteristics. During the earlier part of the season, WDs are much less effective to steer moisture towards the UIB than in the later part, as shown by the smaller positive anomaly of moisture flux intensity (Fig 4.21). Furthermore, the shape of the jet when precipitation occurs (Fig 4.20-B and C) is also similar to the low (high) $Q700$ case in the earlier (later) part of the season (Figure 4.17-B and E). In fact, similarly to the low (high) $Q700$ composite, WDs occurring during the earlier (later) part of the season, drive stronger rear dry advection (front moist advection) leading to a strengthening (weakening) of the trade winds and a weakening (strengthening) of the westerly moisture pathway (not shown).

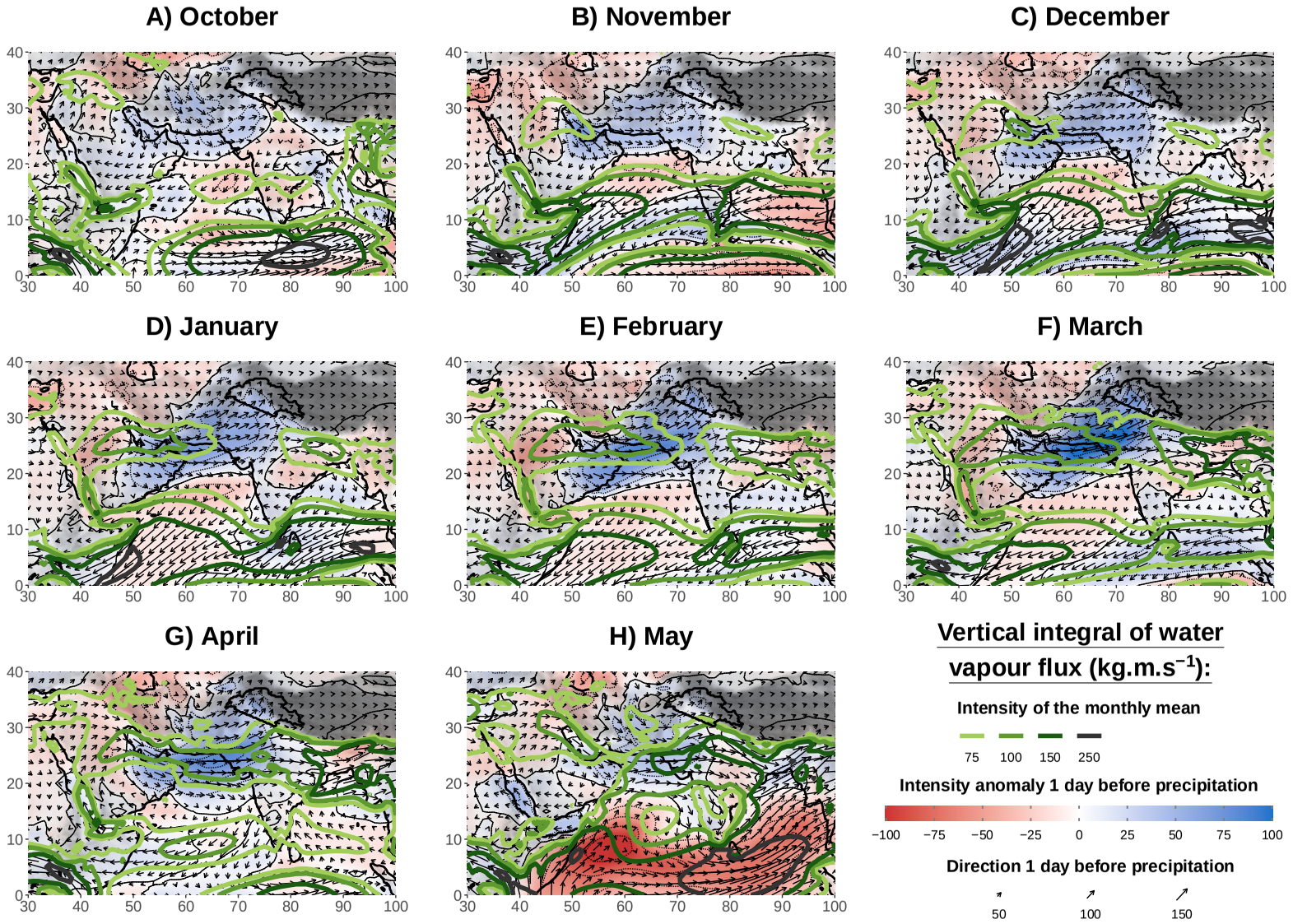


Fig. 4.21 Maps of vertical integral of water vapour flux for each winter month. The thick green contour lines represent the monthly mean intensity. The arrows are one day lead composites based on the same selection as in Figure 4.18 (C to J). The colours are the composite intensity anomaly compared to the monthly mean (thin contour lines every 25 kg m s^{-1}. Non-significant anomalies at the level 95% are shown in white.

To summarise, the SWJ is particularly important to explain the seasonality of WDs. First, its position enables the propagation of WDs to the UIB, determining the WD active season. Second, its strength determines the potential strength of the altitude low characterising a WD and explains the seasonality of WDs. Thirdly, the balance between the Arabian Jet and the East Asian Jet, the two parts of the SWJ in the area, changes the low level circulation associated with the WDs, both in term of cross-barrier winds in the UIB and larger scale moisture transport. This effect explains the wetter spring than autumn.

4.8 Summary

This study has developed an in-depth analysis of the synoptic variability of precipitation in the Upper Indus Basin (UIB) during winter using PC regressions and composites. The atmospheric circulation related to precipitation and particularly Western Disturbances (WDs) have been discussed. Precipitation intensity is explained by various WDs characteristics. Several processes explaining WD initiation, growth, decay, and interaction with the relief are suggested. These findings offer new insights on the precipitation and WD seasonality.

Figure 4.22 summarises the different atmospheric features and processes discussed. First, precipitation events are related to an area of anticyclonic anomaly over Eastern Europe and western Russia with a maximum intensity four days before the precipitation event (Figure 4.22, 1). This anomaly is likely the result of a Rossby wave propagating in the Atlantic Jet stream and forms a ridge from the Atlantic subtropical high. It is enabled by a stronger than usual positive NAO and tilted jet. The anomaly can directly trigger a wave in the Arabian Jet over the eastern Mediterranean Sea through a Rossby wave breaking. Alternatively, the anomaly fuels the development of a pre-existing WD precursor by driving cold air to the rear of the WD and increasing baroclinicity (5). The presence of a trough east of the anticyclone anomaly squeezes the cold air into a corridor; the advection enhances as the WD approaches the trough and interacts with it. The resulting cold anomaly is also in part responsible for the stronger Arabian Jet associated with WDs (2). The cold air advection has an ambivalent role as it is accompanied by dry air that can suppress precipitation (9).

Winter moisture transport is driven by the Arabian subtropical high. North of it, a westerly moisture pathway connects the Red Sea with the Persian Gulf and the north of the Arabian Sea, while trade winds blow over the rest of the Arabian Sea. The passing of a WD enhances the westerly moisture pathway (3), while trade winds weaken it (4). As the WD approaches the UIB, it steers the moisture across the Indus Plain into the notch formed by the relief. The presence of an anticyclonic anomaly east of the WD further enhances the meridional transport (6).

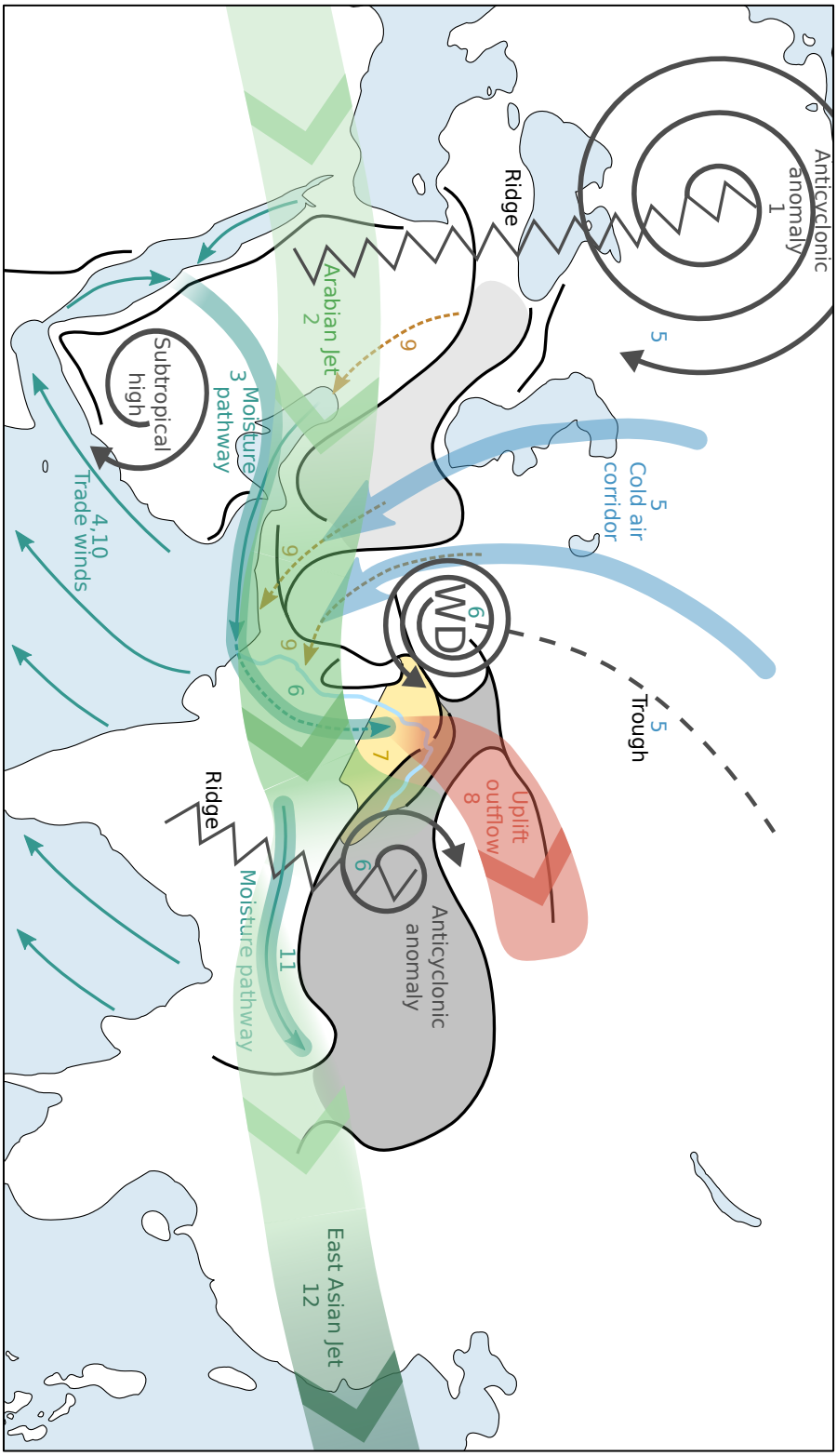


Fig. 4.22 Diagram summarising the synoptic atmospheric circulation associated with the passing of a WD.

Eventually, the moisture converges at low-level (around 700 hPa) in the UIB, over the foothills, and the uplift triggers condensation and precipitation (7). It also allows a mixing and a warming of the air column in the UIB, which counterbalances the cold air advection associated with the WD (cf. 5). The resulting decreased temperature gradient in the UIB allows the upper-level cyclonic circulation of the WD to further propagate downward and enhance the cross-barrier moisture transport. This positive feedback is suggested as a first process explaining WD interaction with the relief in the UIB. The uplift also advects low potential vorticity values to high altitude (250hPa) as indicated by the swift anticyclonic turn of its outflow (8). The presence of lower potential vorticity increases the zonal gradient of potential vorticity, which increases the baroclinic circulation. The uplift is enhanced in the UIB as it can be inferred by the break in the Arabian Jet, helping precipitation. Meanwhile, downward velocities to the east stretches and strengthens the WD cyclonic circulation. This baroclinic interaction is suggested as the second feedback process explaining WD interaction with the relief in the UIB. However, the positive interaction is short-lived since, since, when the WD reaches the UIB, the interaction reverses and the WD quickly weakens.

In parallel to the negative baroclinic feedback, dry continental air intrudes in the Indus Plain and cuts off the moisture supply to the UIB, which effectively stops the precipitation. The extra remaining moisture is then pushed towards Northeast India, while trade winds in the Arabian Sea re-intensify. Finally, as the result of the WD interaction with the relief, the Subtropical Westerly Jet weakens to the east of the UIB and the anomaly propagates into the East Asian Jet two days later.

This idealised scenario hides large variability in the different features discussed in Figure 4.22, leading to varying precipitation intensity. As expected, the intensity of the upper level disturbances impacts low-level cross-barrier winds. However, this study suggests that the thermal structure of the WD is equally important: higher baroclinicity increases the low-level circulation and the hot and moist southerly advection ahead of the WD. Finally, moisture supply is largely affected by the development of the WD. When the WD develops from the cold air advection, its interaction with the moisture pathway is limited and it rather drives dry air into the Indus Plain. By contrast, when the WD propagates closer to the Arabian Jet, its intensity is more limited but the stronger baroclinic interaction leads to a proportionally stronger low-level circulation. Meanwhile, its southern position leads to a stronger impact on the moisture pathway.

Furthermore, these different findings help to understand the precipitation seasonality. Particularly, the precipitation peak is delayed by almost two months compared to the WD peak activity. One month is explained by the moisture content seasonality, while the remaining difference is related to changes in baroclinicity. Precipitation seasonality is closely related to

that of the Subtropical Westerly Jet. The winter jet position enables the WDs to reach the UIB, the jet strength relates to WD intensity, and the balance between the Arabian Jet and the East Asian Jet impacts the WD dynamic in the UIB. Particularly, that balance explains the wetter spring than autumn.

Finally, two non-linear relationships are observed that link precipitation to WD intensity. First, the relationship between moisture transport at 700 hPa and precipitation is better characterised by a quadratic regression, due to varying lower level moisture transport. Second, the relationship between the cross-barrier winds at 700 hPa ($W700$) and the WD intensity (defined as geopotential height anomaly at 300hPa) is also non-linear for higher values of $W700$. This behaviour is likely related to the baroclinic interaction with the relief. Consequently, a small increase in the SWJ, which relates to the WD intensity, may result in a large precipitation change, and could be used as a signal to investigate seasonal variability.

The findings in this study offer different avenues for further work. First, they set up the atmospheric circulation to investigate meso-scale phenomena that affect precipitation patterns, and particularly whether precipitation occurs over the plain or at high altitude. Second, the processes described here can improve the understanding of teleconnections as it was done for NAO. Finally, the representation of those processes and particularly of the Subtropical Westerly Jet are to be tested in climate models.

Chapter 5

Representation of winter precipitation in climate simulations: origin of biases and variability across scales

5.1 Introduction

The previous chapters have selected datasets (Chapter 2), designed tools (Chapter 3), and discussed atmospheric features (Chapter 4) with the aim of evaluating the use of precipitation output in climate simulations. The actual evaluation is conducted in this chapter, with a focus on winter precipitation in the Upper Indus River Basin (UIB) as in Chapters 3 and 4. The evaluation includes the representation of the amount and seasonality of the precipitation, but also its variability from the inter-annual to millennial time scale.

A number of previous studies have investigated winter precipitation biases in global climate models in the UIB or neighbouring areas (e.g. Ahmed et al., 2019; Hunt et al., 2019; Meher et al., 2017; Palazzi et al., 2015, 2013). These studies have focused on the simulations produced for the Coupled Model Intercomparison Project phase 5 (CMIP5; Taylor et al., 2012). They all found that climate simulations overestimate winter precipitation with respect to rain gauge-based datasets, although the spread between simulations is large (Palazzi et al., 2015). Yet, as discussed in Chapter 2, rain gauge-based datasets tend to underestimate precipitation, which reduces the confidence in this finding. Particularly, some simulations are drier than the ERA-Interim dataset (Hunt et al., 2019; Palazzi et al., 2013) and might better estimate precipitation amount than the authors have suggested. Palazzi et al. (2015) have also investigated the seasonality of precipitation and revealed the problematic diversity of the simulated seasonal cycle. The authors suggest that model resolution and the implementation

of sulfate aerosol effects may partly explain this behaviour. The findings on the drivers of winter precipitation in Chapter 4 is re-employed here to further discuss the origin of the average and seasonal biases in the most recent CMIP6 climate simulations.

Various teleconnections affecting UIB precipitation variability on an inter-annual timescale have been exposed in the literature. For example, the North Atlantic oscillation (NAO) affects precipitation through a modulation of the subtropical westerly jet (Filippi et al., 2014), which has been further clarified in Chapter 4. The relationship between UIB precipitation and the El Nino-Southern oscillation (ENSO) is also strong (Cannon et al., 2016b; Dimri, 2013b) and can be explained by the stronger dynamics of Western Disturbances (WDs; Yadav et al., 2013, 2010) and an increased moisture supply (Mariotti, 2007; Syed et al., 2010) during El Nino events. Interestingly, the strength of both of these teleconnections exhibit decadal variability in the historical record (Filippi et al., 2014; Yadav et al., 2009), while ENSO and NAO signals have varied on various timescales across the Holocene (Hernández et al., 2020). Hence, the temporal modulation of the teleconnections has an impact on the longer timescale variability of the precipitation. A good representation of the teleconnections in climate simulations is therefore a prerequisite for the analysis of the precipitation variability on longer timescales. Most climate models in CMIP5 and the most recent CMIP6 simulate the ENSO variability and the positive impact of El Nino events on winter precipitation in the UIB (Brown et al., 2020). However, a more detailed analysis of atmospheric and oceanic global teleconnections pattern in relation to precipitation in the UIB is needed to further validate climate simulations.

On millennial time scales, orbital forcing becomes the predominant driver of climate variability. Since the mid-Holocene, 6000 years before present, the precession has changed phase while the obliquity decreased (Berger and Loutre, 1991). The change in precession affects the inter-hemispheric heat repartitioning: during the mid-Holocene, the Earth's perihelion (i.e. the position on the orbit the closest to the sun) coincided with boreal summer, which led to a strengthening of the boreal monsoon (Kathayat et al., 2016; Liu et al., 2004). By contrast, present day Earth's perihelion coincides with austral summer, which leads to a strengthening of the austral monsoon. In comparison, little is known about how the winter season is affected by this change in seasonality due to a lack of proxy data close to the UIB (Dixit et al., 2018; Giesche, 2020; Giesche et al., 2019) or elsewhere (e.g. Perșoiu et al., 2017). Model data hypothesised a weakening or northward shift of the Subtropical Westerly Jet (Hou et al., 2017; Hunt and Turner, 2019). The obliquity also impacts the temperature gradient between the equator and the higher latitudes, but only on average over the whole year (Mantsis et al., 2011; Wu et al., 2016). One possible impact of the stronger obliquity during the mid-Holocene and the lower insolation near the equator is a weaker ENSO variability,

which is both found in proxies (Emile-Geay and Tingley, 2016) and in climate simulations (Chen et al., 2019). These correlations highlight again the importance of teleconnections even at the millennial timescales.

This chapter is split into three main sections. The first describes the model and observational datasets used, the second investigates the seasonal biases of winter precipitation in climate simulations and its change since the mid-Holocene, and the third discusses the representation of winter precipitation variability across timescales.

5.2 Data

5.2.1 Observations and reanalyses

As in Chapter 3 and 4, ERA5 reanalysis (Hersbach et al., 2020) is used here. The analysis presented in Chapter 2 (also in Baudouin et al., 2020b) has shown that ERA5 provides precipitation estimates closest to rain gauge-datasets among various reanalysis datasets. In addition, it provides 3D sub-daily estimates of geopotential, wind and specific humidity at 0.25° resolution, between 1979 and 2018. The GPCC rain gauge-based dataset is used as a complement here (Schneider et al., 2018). Chapter 2 also showed that GPCC is one of the most accurate observational datasets for the area, and most importantly provides data since 1891, which proves useful for lower frequency variability. Note, however, that its quality before 1950 remains unchecked, due to a lack of independent datasets for validation (Chapter 2; Baudouin et al., 2020b). Observed sea surface temperatures (SST) are from the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) v1.1 for the period 1891-2018, to match the time period of GPCC and ERA5 (Rayner et al., 2003). HadISST is also used to compute an ENSO index, the Oceanic Niño Index (ONI; Trenberth, 1997). This ENSO index is defined as the 3-month average of SST anomalies in the Niño3.4 region (5°N-5°S, 170°-120°W). Finally, an observation-based NAO index is used (Hurrell et al., 2013)[1]. This index is defined as the first principal component of sea level pressure (SLP) over the North Atlantic region (20°-80°N, 90°W-40°E).

5.2.2 IPSL climate simulations

In order to investigate the biases of climate models in more depth, the analysis is limited to simulations with Earth system models from the IPSL family. Table 5.1 summarises the characteristics of the model data that has been used. The historical, pre-industrial control

[1]NAO Index Data provided by the Climate Analysis Section, NCAR, Boulder, USA. Accessed 05-08-2020

Table 5.1 Characteristic of the model data used in the analysis.

Model	Experiment	Variant	Time period (Model years)	Time resolution	Level	Variables
IPSL-CM6A-LR	Historical	r1i1p1f1	1980-2005	3-hourly	surface	precipitation
	piControl	r1i1p1f1	2000 years (1850-3049)	3-hourly	surface	precipitation
				monthly	surface	SST, SLP
			670 years (1870-2539)	6-hourly	700 hPa	wind, specific humidity
					250 hPa	wind
				monthly	surface	precipitable water, vertically integrated moisture transport
	mid-Holocene	r1i1p1f4	300 years	monthly	surface	precipitation, SST, SLP, precipitable water, vertically integrated moisture transport
IPSL dynamic vegetation	Transient Holocene (6ka BP to 1950 CE)	Not specified	6000 years	monthly	surface	precipitation, SST

(piControl), and mid-Holocene experiments were produced for CMIP6 (Eyring et al., 2016) with the IPSL-CM6A-LR model (Boucher et al., 2020) and are investigated here. The historical experiment has already been used in Chapter 3 to validate the relationship between cross-barrier moisture transport and precipitation (Baudouin et al., 2020a). Here, it is used to compare the precipitation with that of the piControl, and it is also used to assess the impact of different greenhouse gases (GHG; i.e. carbon dioxide, methane and nitrous oxide) concentrations (i.e. varying GHG concentrations from 1850 to 2005 are used in the historical run, and constant 1850 concentrations in the piControl, while the orbital conditions are identical and correspond to today's conditions). The piControl experiment is preferred for most of the analysis due to its larger data availability. Particularly, the variant r1i1f1p1 offers a timeseries of 2000 years for precipitation, SLP, and SST, which allows for investigation of lower frequency variability. A shorter time period of 670 years is available for sub-daily 3D winds and specific humidity.

The mid-Holocene experimental setup has been designed using the Paleoclimate Modelling Intercomparison Project - Phase 4 protocol (PMIP4; Otto-Bliesner et al., 2017). However, the variant r1i1p1f4, selected in this study for its longer time period (300 years), uses 3D mineral dust forcing provided by Samuel Albani instead of the 1850 pre-industrial values. Unfortunately, wind and specific humidity are not available on standard pressure levels at a daily or sub-daily time resolution. Thus, the relationship between precipitation and moisture transport at 700 hPa (cf. Chapter 3 and 4) cannot be analysed for this experiment.

Finally, a transient simulation of the last 6000 years, until 1950 CE, is investigated (Braconnot et al., 2019). This simulation was performed with a modified version of the IPSL-CM5A-MR that was used for CMIP5. Changes were made to the land surface and hydrology components, in particular with the addition of a dynamic vegetation module. The atmospheric component uses the same grid as IPSL-CM6A-LR so that the comparisons do not need interpolation. The two transient external forcings are orbital parameters and GHG concentrations while aerosols (dust and sea salt) are fixed at 1860 level.

5.3 Seasonality

5.3.1 Method

Winter precipitation is defined here as it was in Chapter 4: the precipitation that falls from October to May, while the UIB area is the same as in the Chapters 2, 3, and 4, although it was adapted to the IPSL resolution (cf. Chapter 3, Figure 3.8). Wind and specific humidity at 700 hPa are available at sub-daily (6-hourly) resolution for the piControl experiment,

so that the relationship between wind, specific humidity, and precipitation in the UIB can be investigated using the same technique used in Chapters 3 and 4 (i.e. a PC regression; Baudouin et al., 2020a)

To get comparable results between ERA5 and the climate model output, ERA5 data are first coarsened to the model's spatial and temporal resolution. Then, the same PC regression as in Chapter 4 is performed. To recall, precipitation is predicted using the product of wind and specific humidity (hereafter moisture transport) at 700 hPa at the start of the precipitation accumulation period. The regression is optimised for non-negative predicted values, while insuring an unbiased result. The predictors are the first principal components of moisture transport around the UIB. The first 15 principal components are selected in this chapter, which is lower than the number used in Chapter 4 (46), due to the coarser spatial resolution. Figure 5.1 presents the pattern of the regression coefficients. The result is similar to Figure 4.1, confirming that the difference in resolution does not impact the main results. The coefficient of determination (R2) is 0.844, which is slightly lower than previously (0.878; Section 4.3), due to the lower spatial resolution of the predictors (i.e. fewer principal components).

A similar PC regression was employed with IPSL-CM6 historical simulation in 3.4.5 (Baudouin et al., 2020a). It showed that, as in the reanalysis dataset, moisture transport along the Himalayan foothills is important for producing precipitation in the simulations. However, using the PC regression on model outputs does not make it possible to determine whether a

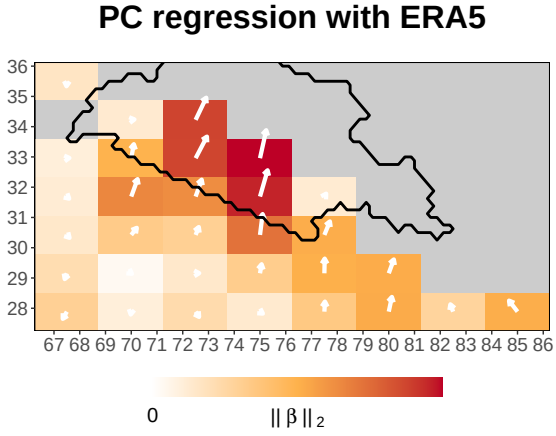


Fig. 5.1 Result of the PC regression of precipitation in the UIB with moisture transport at 700 hPa. Same as Figure 4.1-A, but with the moisture transport coarsened to the IPSL-CM6A-LR spatial resolution, and with a 6-hourly timeseries instead of 3-hourly. The first 15 principal components are selected. The arrows represent the regression coefficients associated with both zonal and meridional direction of moisture transport at each location. The colour is proportional to the arrow's length and represents the overall importance of the location.

change in the precipitation rate is related to an increased moisture transport or an increased efficiency to convert moisture transport into precipitation. So, instead of computing new regression coefficients, the coefficients derived from the fit with ERA5 data are re-employed and applied to the model data. Doing this, the different timeseries (predicted precipitation, moisture and wind contributions) are spatially weighted means using the same weights, so that the various variables can be compared.

5.3.2 Biases in the seasonality of precipitation in the piControl simulation

In Figure 5.2-A, the thick lines show the seasonality of precipitation in the UIB from ERA5 reanalysis (black), historical run (grey), and the piControl run (blue). The differences between the historical and piControl run are small (the historical winter precipitation is 3% higher), and can be related to the smaller time period considered for the historical run (1979-2014) as much as to the change in greenhouse gases concentration. As discussed in Chapter 3, for the historical experiment the bi-modality of the precipitation in the UIB is mostly absent, with summer precipitation heavily underestimated. By contrast, the winter precipitation is substantially overestimated (+46% for the period October to May compared to ERA5). The variations are otherwise well represented, with a distinctive peak in March, although somewhat delayed in the historical run. The overestimation during October-December, the driest period in ERA5, further minimises the weak summer monsoon precipitation peak in August-September.

The dotted lines in Figure 5.2-A represent the seasonality of the predicted precipitation using moisture transport. The mean predicted winter precipitation from the piControl is 34% higher than that of ERA5, which indicates that an enhanced moisture transport is responsible for most, but not all, of the precipitation difference. Indeed, the precipitation derived from moisture transport is 8% lower than the precipitation simulated by the model (cf. blue lines), while for ERA5, the prediction is unbiased by definition. The difference is particularly evident during the wettest winter months. This difference suggests that the moisture transport is more efficient at producing precipitation in the IPSL climate model. Explanations for this behaviour range from differences in relief representation due to the resolution, differences in microphysics parametrisations, as well as differences in other environmental conditions (e.g. temperature, humidity, or wind at other altitudes).

Figure 5.2-B explores the seasonality of the timeseries of the contribution from wind and specific humidity (respectively $W700$ and $Q700$, cf. definitions in Chapter 4). Note that only the positive values of $W700$ are considered, as negative values do not impact precipitation

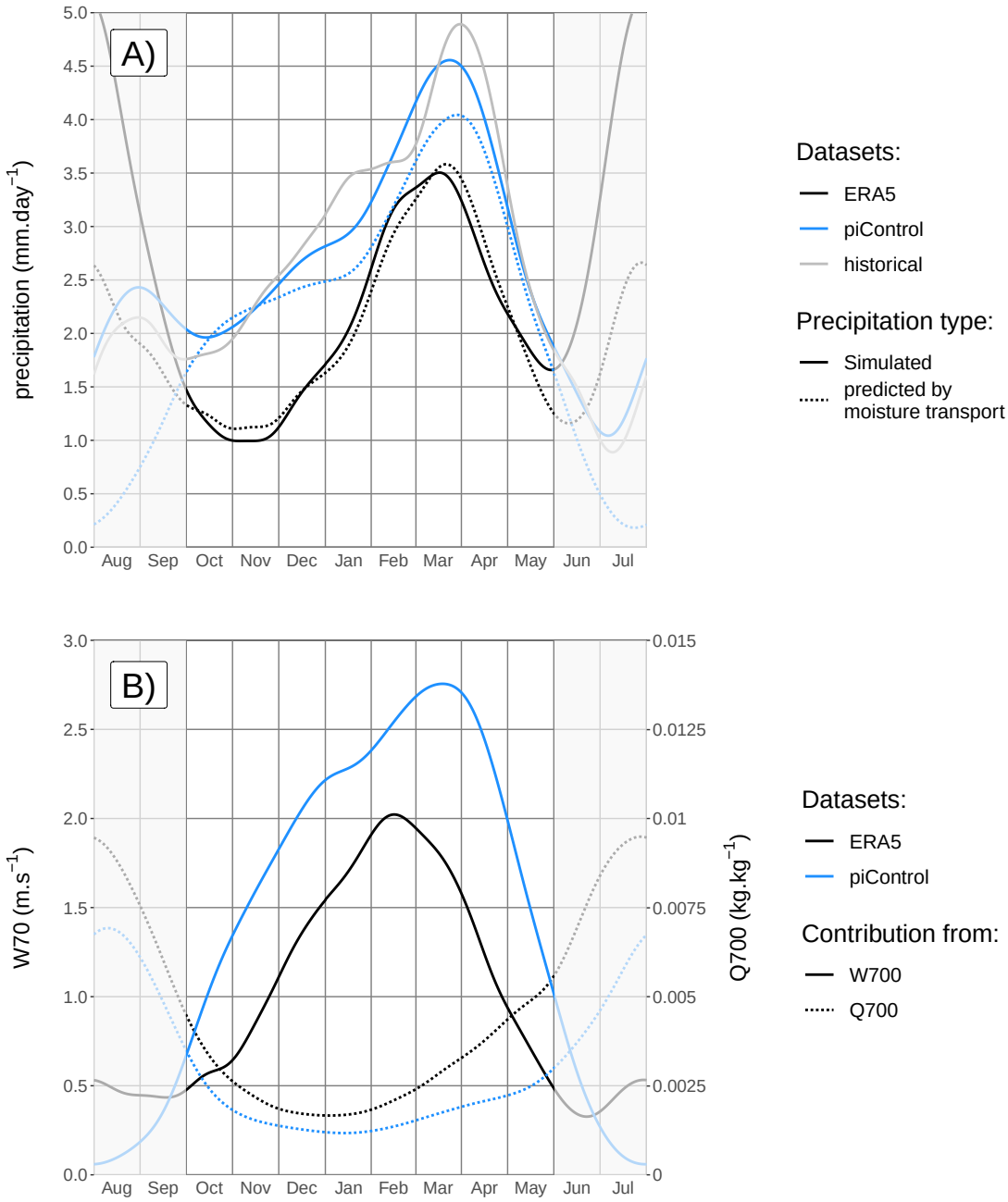


Fig. 5.2 In A, seasonality of precipitation (plain line) in ERA5 (black), the piControl (blue) and historical (grey) simulation using IPSL-CM6-LR. The seasonality of the precipitation predicted by moisture transport is given by the dotted line. This prediction is only performed for ERA5 and piControl. The months June, July, August, and September are greyed out as they are not used to fit the precipitation with moisture transport at 700hPa; a prediction is still computed but results in large biases. In B, the seasonality is given for the two variables composing the moisture transport: specific humidity ($Q700$, dotted line) and wind ($W700$, plain line) for both ERA5 and piControl run. The values of W700 not contributing towards precipitation (i.e. negative values) are fixed to 0.

(i.e. positive $W700$ indicates cross-barrier winds in the UIB). Interestingly, the biases of the two variables have opposite signs: $Q700$ is 38% lower in the piControl simulation than in ERA5, but $W700$ is 58% higher. The difference is relatively similar across the whole winter period (October to May), but tends to increase for the last three months (March to May). Surprisingly, multiplying the respective bias of $W700$ and $Q700$ for the whole season gives a result close to 1 (0.986): that is, if treated independently, the biases from $W700$ and $Q700$ almost exactly cancel each other out. In other words, the seasonal bias of moisture transport in the simulation is related to a stronger spatial and temporal correlation between the wind and specific humidity fields than in ERA5 (cf. Section 3.4.2).

5.3.3 Western Disturbances in the piControl simulation

The origin of the biases of moisture transport, wind, and specific humidity in the piControl simulation is further investigated using the results from Chapter 4. Figure 5.3 shows the seasonal anomaly of geopotential height, averaged over the 10% strongest 6-hourly accumulation periods of precipitation, for each month of the year. The results are presented on a longitudinal cross-section along 31.69°N, i.e. across the Indus Plain, which can be compared to Figure 4.18. For all months between October and June, the highest precipitation events are related to the presence of a minimum anomaly of geopotential height around 300hPa. Further, the minimum anomaly for a given altitude is tilted westward with increasing height (vertical line in Figure 5.3), which indicates the presence of a baroclinic instability. These two elements characterise the WDs discussed in Section 4.4. Consequently, the simulation represents the correct synoptic driver of winter precipitation. Yet, some differences can be seen with respect to Figure 4.18. First, the minimum anomaly is generally located at a lower altitude (below 300 hPa) in the simulation compared to ERA5 (above 300 hPa), which suggests a lower altitude for the dynamic tropopause too. Second, the increased altitude of the minimum anomaly in October and May is not as evident. Finally, the pattern in June is probably the most striking difference, as it shows that, in contrast to ERA5, upper-level disturbances are still present and triggering precipitation in the simulation although no baroclinic processes are involved (cf. absence of westward tilt). This seasonality change can be related to the delayed and weaker summer monsoon precipitation in the simulation.

Since WDs are responsible for precipitation during the months from October to May in the piControl simulation, the atmospheric circulation during precipitation events can be investigated in greater detail. Figure 4.13 cannot be reproduced as vertically integrated moisture transport is not available at a sub-daily resolution for the simulation. Further, the method of the composite used in Figure 4.13 is not suited to compare ERA5 and the piControl simulation, due to the differences in the precipitation mean between the two datasets. Instead

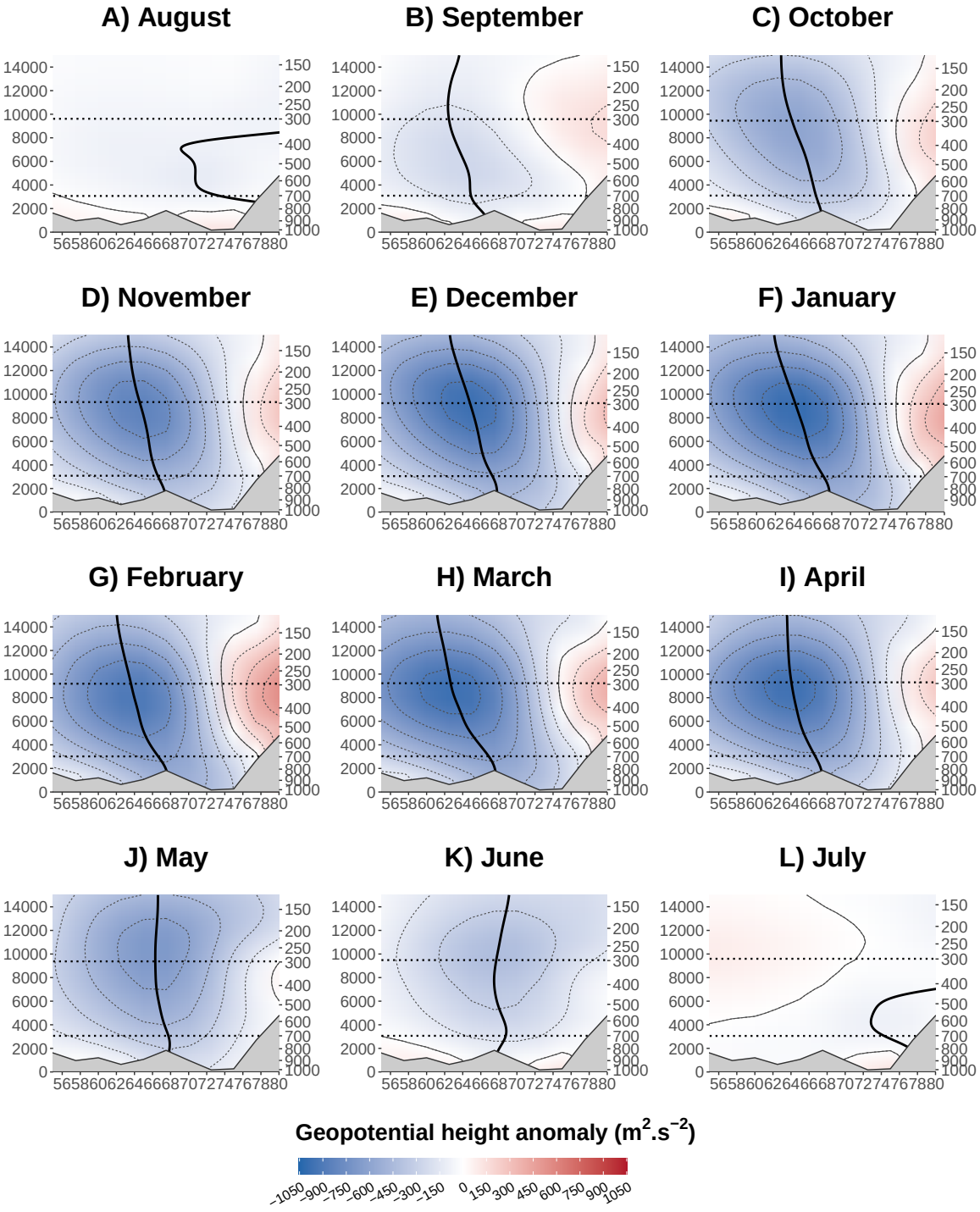


Fig. 5.3 Vertical cross-section at 31.69°N and between 56 and 80°E using piControl data: composite of geopotential height anomaly for the 10% highest 6-hourly precipitation rates, per month. The vertical line indicates the longitude of minimum anomaly for each altitude.

a point-wise regression is performed between precipitation in the UIB, and both absolute wind velocity and the seasonal anomaly of specific humidity at 700 hPa, at different time lags (Figure 5.4).

The pattern for ERA5 (Figure 5.4-A to C) are very similar to Figure 4.15: a positive moisture anomaly develops in the Indus Plain and to the west, supported by south-westerly winds on the day prior to the precipitation event (Figure 5.4-A). This moisture anomaly moves to the upper part of the basin where the cross-barrier winds trigger convergence, producing precipitation (Figure 5.4-B). In the mean time, a negative anomaly of moisture develops to the west of the cyclonic circulation with northerly to westerly continental wind (Figure 5.4-A and B). This moisture anomaly quickly invades the Indus Plain after the precipitation peak, despite cross-barrier winds still present along the Himalayas (Figure 5.4-C).

While a positive moisture anomaly also develops in the piControl simulation east of a cyclonic circulation, other elements are different. First, the moisture anomaly is supported by southerly winds, that originate at a lower latitude than in ERA5 (Figure 5.4 D and E). As a result, the moisture anomaly is almost twice as important. Note that the specific moisture is lower on average in the piControl simulation (cf. Figure 5.2-B, also later Figure 5.6). This bias needs to be compensated by either an enhanced cross-barrier winds in the UIB or an increase in specific moisture anomaly to result in the same amount of precipitation. Figure 5.4 indicates that stronger cross-barrier winds are in fact not needed as the change in wind circulation is able to advect more moisture towards the UIB. Second, despite strong northerly winds to the rear, the negative moisture anomaly is less well pronounced (Figure 5.4-D and E), which may also be related to the average lower moisture content as just discussed. The stronger meridional circulation on both sides of the cyclonic circulation is likely related to the coarser relief representation, that can be seen in 5.3. Indeed, the various ranges between the Iranian Plateau and the Tibetan plateau tend to disrupt the meridional circulation, while the Persian Gulf funnels the zonal flow. Finally, the negative moisture anomaly does not progress to the Indus Plain after the precipitation peak in the simulation (Figure 5.4-F). It is possible that the stronger meridional circulation limits the impact of westerly winds, which is responsible for the drying in the Indus Plain through a Foehn effect down the Iranian Plateau and the Suleiman Range (cf. Section 4.6). Both the stronger moisture anomaly and the absence of dry air intrusion explains the better overlap between areas of wind and moisture anomalies already discussed in the previous section of this chapter (5.3.2), and lead to the overestimation of moisture transport and precipitation in winter.

This analysis of the synoptic circulation does not completely explain the origin of the positive wind overestimation and dry bias in the piControl simulation. While the stronger

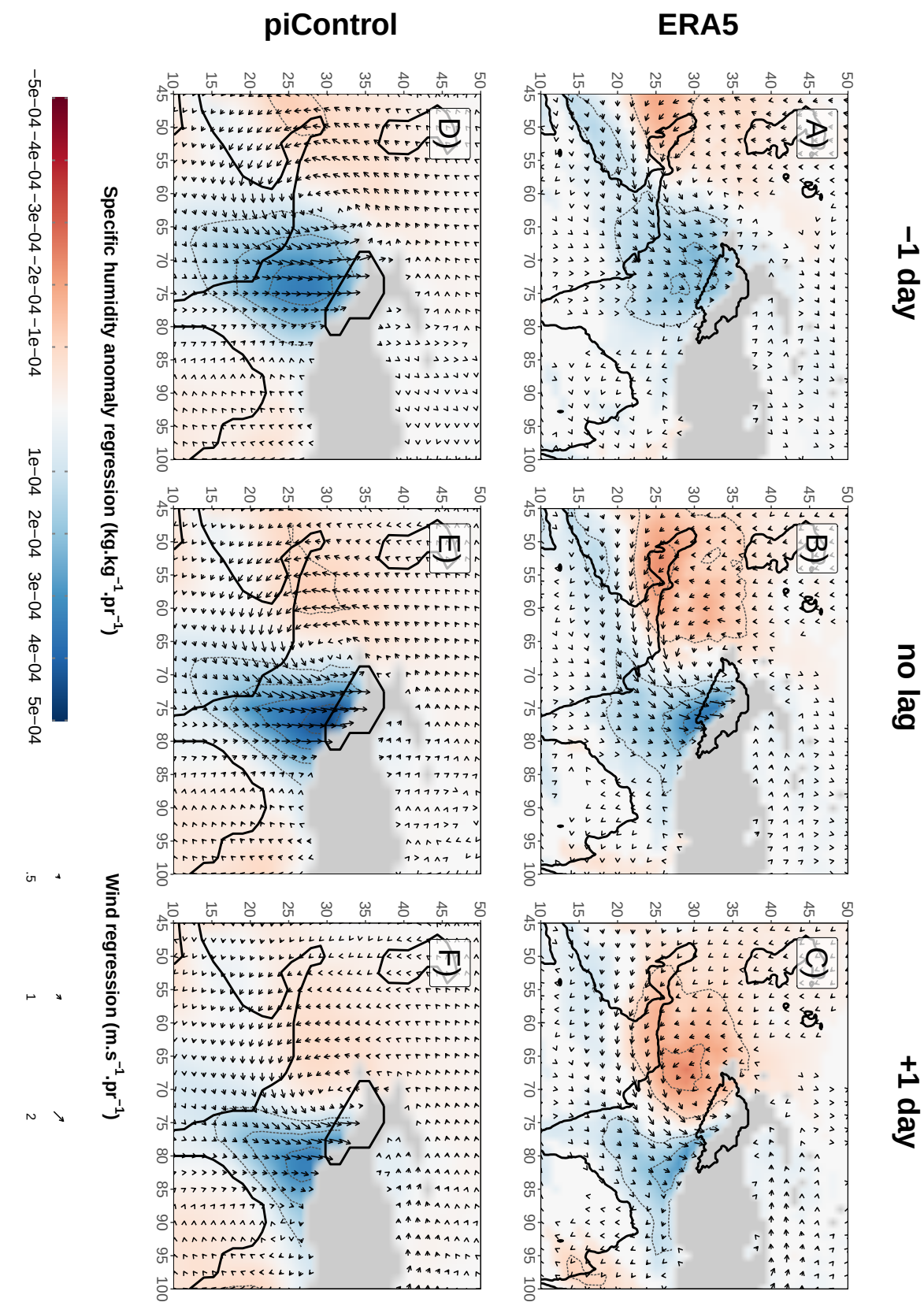

ERA5
piControl
−1 day
no lag
+1 day
A)
B)
C)
D)
E)
F)
Specific humidity anomaly regression (kg.kg^{-1}.pr^{-1})
−5e−04 −4e−04 −3e−04 −2e−04 −1e−04
1e−04 2e−04 3e−04 4e−04 5e−04
Wind regression (m.s^{-1}.pr^{-1})
.5
1
2

Fig. 5.4 (Previous page) Regression maps of precipitation in the UIB (indicated by the black contour line) with absolute wind (arrows) and seasonal anomaly of specific humidity at 700 hPa (colours). The regression coefficients are not standardised, so that the values represent the anomaly of moisture transport and the wind field associated with a precipitation event of 1 mm in 6 hours. The regression is performed for two datasets: ERA5 (A to C) and piControl (D to F), as well as for different lags of specific humidity and wind: one day before the precipitation event (A and D), at the start (B and E), and one day after (C and F). The surface lying above 700 hPa (at minimum pressure) is represented in grey. When the regression coefficients of the two wind directions are not significant at the level 0.05, the arrow is not plotted. Similarly, when the regression coefficient of specific humidity is not significant, it is fixed to 0. The dotted lines are iso-contours of specific humidity regression coefficients as indicated by the labels in the legend.

---

meridional circulation can explain part of the wind overestimation, another reason can be found in the representation of the subtropical westerly jet.

Figure 5.5-A shows the seasonality of the wind speed at 250 hPa at various latitudes along 70°E. It presents both the mean wind speed in the simulation (green contour lines) and the anomaly compared to ERA5 (colours). The jet stream in winter is more than $5\,\mathrm{m\,s^{-1}}$ stronger in the piControl simulation, which corresponds to a 10 to 20% increase in wind speed. In contrast, the jet completely disappears in summer, instead of moving northward (cf. 4.20). The winter increase can be explained by the weakening of the polar jet that moves between the subtropical jet and the Arctic (see the negative wind speed anomaly at these latitudes in Figure 5.5-A). This way, the kinetic energy is transferred from the polar jet to the subtropical jet. Hunt et al. (2019) suggested that the coarse model resolution could be responsible for this difference, but biases in heat sources and sinks over the Tibetan Plateau can also be responsible for the biases in subtropical jet. Indeed, Schiemann et al. (2009) showed that heat sinks over the Tibetan Plateau in winter, along with the surface drag locks the jet to the south of the Plateau. On the contrary, the Tibetan Plateau becomes a major heat source in late spring and throughout summer, which explains the late spring weakening and eventually northward movement of the jet (cf. Section 4.7). Figure 5.5-A shows a larger wind speed anomaly between March and May (above $10\,\mathrm{m\,s^{-1}}$) which effectively maintains the jet intensity for a longer period into spring. This extension explains why WDs remain active up to June in the piControl simulation (cf. Figure 5.3-K). This wind speed bias is likely caused by a delay in the establishment of the heat sources over the Tibetan Plateau. For example, stronger winter precipitation in the UIB, that falls as snow at altitude, may extend onto the Tibetan Plateau itself and play a role in theses biases, through an increased albedo in spring.

Further analyses are needed to prove these hypotheses. Finally, the delay and bias in Tibetan Plateau heat sources play a role in the weaker summer monsoon.

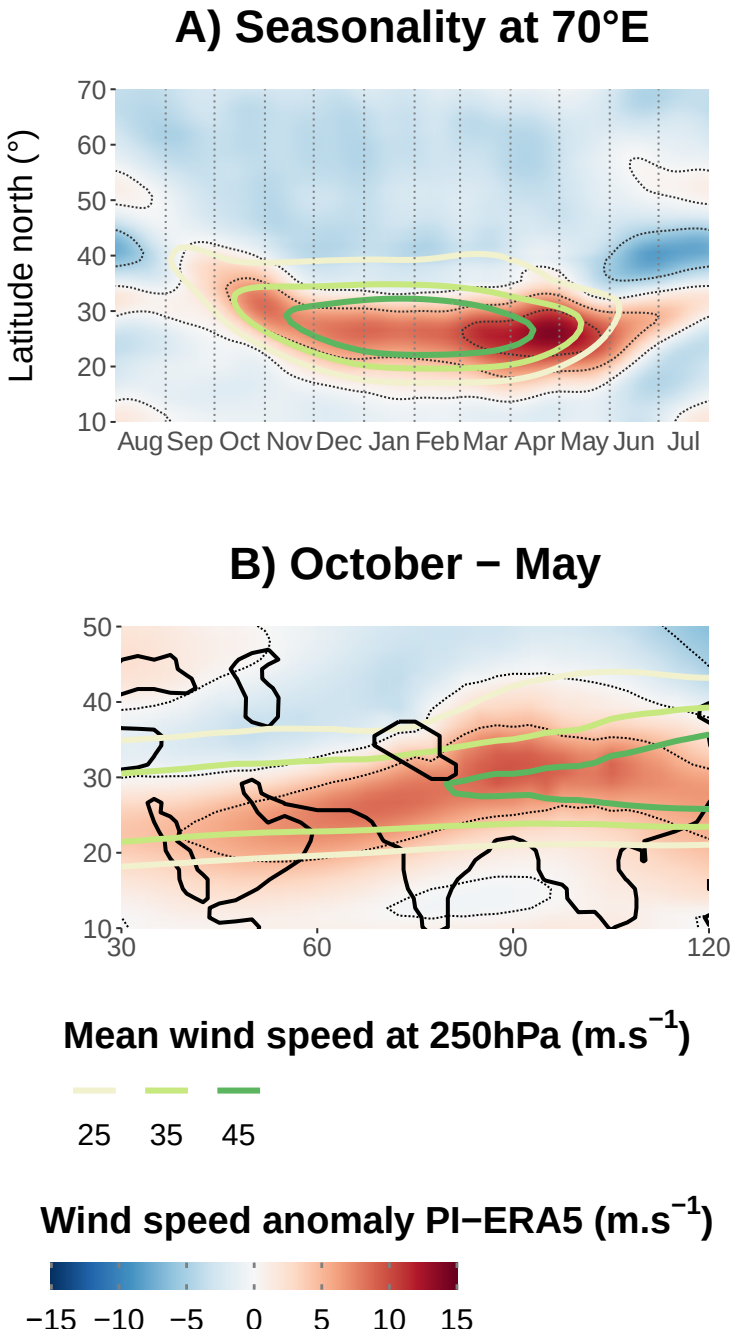


Fig. 5.5 In A, Hovmöller representation of the seasonality of wind speed at 250 hPa and along 70°E. In B, map of the same variable averaged between October and May. The green contour lines represent the wind speed, while the colour shading is the anomaly in the piControl run compared to ERA5.

As discussed in Section 4.5, a stronger subtropical westerly jet in winter fuels deeper WDs, which eventually trigger higher low level cross-barrier winds in the UIB. Therefore, the overestimated wind in the UIB can be explained by the anomaly in the jet speed. Yet, Figure 5.5-B also suggests that this relationship is partially offset by a slightly different configuration of the jet-stream. Figure 5.5-B shows the spatial distribution of mean wind speed at 250 hPa in the piControl simulation and the wind speed anomaly compared to ERA5 during winter. The maximum wind speed anomaly is located just south of the Tibetan Plateau, which further supports the hypothesis that the Tibetan Plateau is responsible for the bias. In contrast, the part of the jet over the Arabian Peninsula is not as anomalously strong compared to ERA5. In fact, when looking at the averages per month (not shown), the Arabian jet never detaches from the stronger East Asian Jet (cf. Figure 4.20 for ERA5). As a result of these anomalies, the UIB tends to remain in the left entrance of the East Asian Jet with increased baroclinic sinking, which limits lower level cross-barrier winds as discussed in Section 4.5.

Finally, the moisture anomaly is investigated using precipitable water and vertically integrated water vapour flux averaged over the period October to May. The piControl simulation is compared to ERA5 in Figure 5.6. Most strikingly, precipitable water is underestimated for the whole domain, except for a few small patches. The underestimation is particularly important over the Indus and Ganges River basins and in Northeast India. This pattern suggests stronger north-westerly winds that accompany the air sinking from the mountain ranges to the west and the north. Again, the anomalous jet position may be responsible. A second area of strong negative moisture bias is the south of the Red Sea. Figure 4.16 showed that strong evaporation occurs over there, and it is likely that weaker evaporation happens in the piControl simulation, potentially caused by weaker surface winds. This bias over the Red Sea is also important to explain the dry bias over the Persian Gulf and towards the Indus Plain, since Section 4.6 showed that moisture is transported from the Red Sea to these areas. In fact, the westerly moisture pathway from the Red Sea to the Indus Plain discussed in Section 4.6 and indicated in Figure 5.6 is much weaker in the piControl simulation and is not any more a distinct area of maximum moisture transport. Instead, moisture transport from the Mediterranean Sea towards Central Asia is enhanced due to an enhanced zonal flow. The weakening of the westerly moisture pathway to the UIB and the more southerly origin of the winds when a WD passes (cf. Figure 5.4-E) suggests that moisture comes mainly from the Arabian Sea in the piControl simulation. Meanwhile, much stronger moisture transport occurs in the trade wind region, off the African coast (Figure 5.6). Given the lower column moisture content, this implies stronger trade winds. This enhanced southward moisture transport is likely the main explanation for the lower precipitable water between 10 and 35°N.

5.3.4 Comparing the other datasets

Figure 5.7 presents the seasonality of precipitation in the UIB for various datasets. For the present era, ERA5 reanalysis and rain gauge-based GPCC are considered for precipitation. Chapter 2 demonstrated that these two datasets were among the most reliable. Yet their differences highlight the uncertainty in the observed amount of precipitation as discussed in Chapter 2. For the IPSL-CM6A-LR model (dashed line) two simulations are considered, the piControl (blue) and the mid-Holocene (red). The seasonality is also presented for the IPSL Holocene transient simulation (dotted line) but for two sub-periods, the first 300 years that

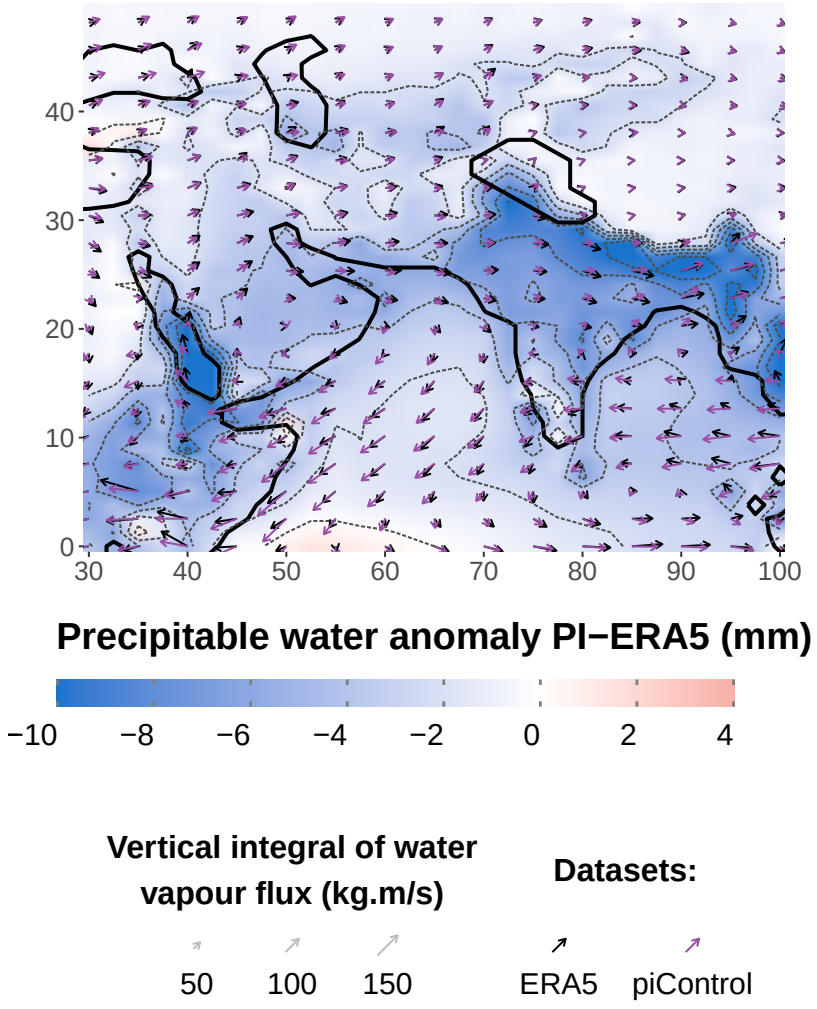


Fig. 5.6 Map of precipitable water anomaly (colour) and total column moisture transport (arrows) in the piControl simulation compared to ERA5, as an average between October and May.

can be compared to the mid-Holocene simulation (blue) and the last 300 years that can be compared to the piControl simulation (red).

Strikingly, the transient simulation is significantly drier than both IPSL-CM6A-LR simulations, despite the strong similarities between the two models. More specifically, the seasonality is reduced with a much weaker peak in March or April. Without access to atmospheric variables such as wind and specific humidity at 700 hPa or precipitable water for the transient simulation, the origin of these differences was not further investigated.

The changes between the mid-Holocene (in red in Figure 5.7) and the pre-industrial conditions (in blue) are investigated by comparing the mid-Holocene and piControl simulations (dashed lines), as well as by comparing the first and last 300 years in the transient simulation (dotted lines). Despite the mismatch in seasonality between the IPSL-CM6A-LR simulations and the transient simulation, some qualitatively robust changes are evident. The largest differences occur in August and the neighbouring months of July and September. Both

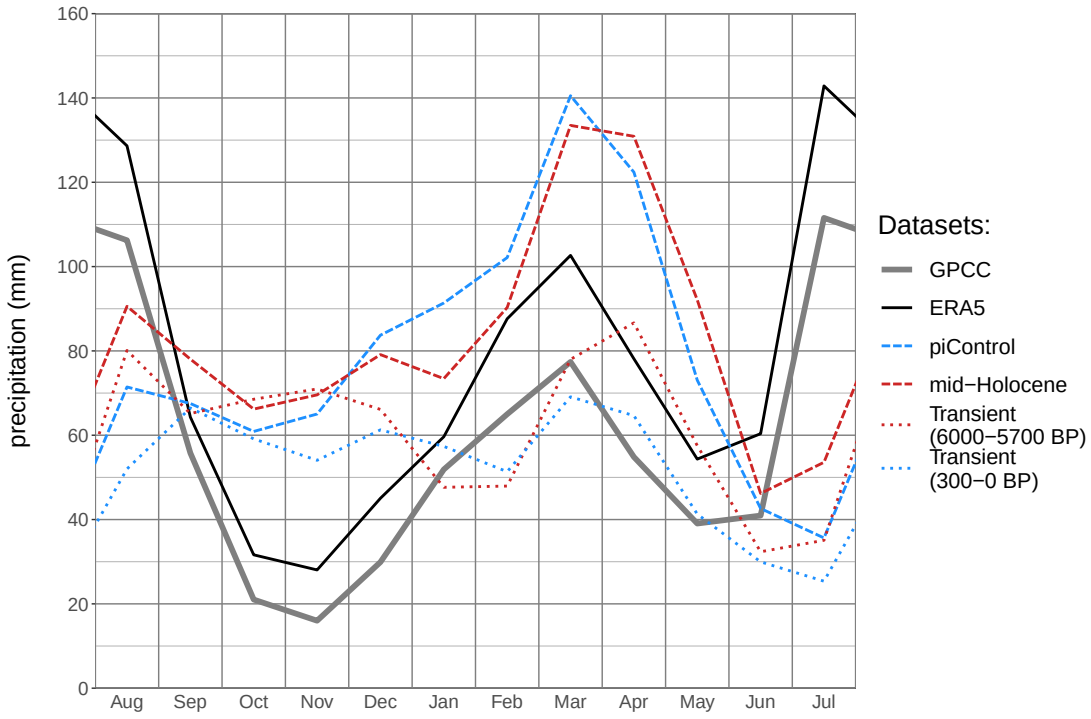


Fig. 5.7 Seasonality of precipitation in various datasets: GPCC (grey, rain-gauge based), ERA5 (black, reanalysis), piControl (blue) and mid-Holocene (red) simulation based on IPSL-CM6-LR, and 6k years transient Holocene simulation. For the latter, the seasonality is given over two sub-periods: the first 300 years (red) and the last 300 years (blue) of the simulation.

IPSL-CM6A-LR and the transient simulation suggest a stronger summer monsoon during the mid-Holocene, which has been documented in other climate simulations (Bosmans et al., 2012; Jiang et al., 2014; Zhao et al., 2005) as well as in climate records (Dixit et al., 2018; Fleitmann et al., 2003; Kathayat et al., 2017), and suggest that the Indus River basin was particularly impacted by this change.

In winter, the changes are more complex. When considering the whole period from October to May, the change is less than 1% for the IPSL-CM6A-LR simulations, but the transient simulation shows a negative trend of 13% across the 6000 years. However, investigating smaller periods shows that the two sets of simulations only disagree on the sign for two months (December and March) and that trends of different signs coexist the rest of the season. In October-November as well as in April-May, the simulations of the mid-Holocene are wetter than the pre-industrial conditions. By contrast, for January-February, the pre-industrial period is wetter. Analysis of the precipitable water for the IPSL-CM6A-LR simulations shows a positive anomaly for the mid-Holocene in October-November likely due to the remnant of moisture of the stronger summer monsoon as well as higher local evaporation. By contrast, lower precipitable water is evident in January-February although the reasons for this pattern are less clear. Hunt and Turner (2019) discussed how the reduced strength of the jet in the mid-Holocene might have impacted intensity of the WDs and thus precipitation during wintertime. However, analysis of the wind speed at 250 hPa in IPSL-CM6 simulations does not support this hypothesis. Another possibility is the lower winter temperature in the mid-Holocene simulations, which is linked to weaker insolation and lower greenhouse gas concentrations compared to the piControl simulation (Brierley et al., 2020). Finally, the origin of the change in precipitation in April-May is even less clear as neither wind nor moisture seem to be different on average between the two periods. Daily or sub-daily wind and moisture fields would be needed to perform a more in depth analysis and to help understanding the origins of the precipitation differences.

As discussed in Section 1.4.1, there is no winter UIB precipitation proxy available to discuss Holocene trends. There is evidence for a wetter early to mid-Holocene in the Levant (Cheng et al., 2015, and references therein) and up to the Zagros mountains in Iran (e.g. Andrews et al., 2020), a region where precipitation is mostly brought by WDs. However, the moisture there has a different origin, the Mediterranean Sea. In fact, previous modelling studies have suggested an opposite reaction to precession changes in the Middle-East compared to the IUB. So, extrapolation of the Holocene trends in the Middle-East to the UIB is not advised.

5.4 Variability and teleconnections

5.4.1 Variability across timescales

5.4.1.1 Haar fluctuations

The variability of the precipitation signal at different time scales is investigated using Haar fluctuations (Lovejoy and Schertzer, 2012). Haar wavelets can be defined as equal to 1 between $-\frac{s}{2}$ and 0, -1 between 0 and $\frac{s}{2}$ and 0 otherwise, where s is a positive number, in unit of time, defining the timescale investigated (cf. the x-axis "period" in Figure 5.8). A Haar fluctuation is obtained by integrating the product of a Haar wavelet with the timeseries investigated T and scaling the result by $\frac{s}{2}$. The fluctuation is dependant on the timescale s considered but also on the location x on the timeseries (i.e. where the centre of the wavelet is located on the timeseries). The fluctuation also simply equals the absolute difference between the average values of the timeseries between $x-\frac{s}{2}$ and x on one hand, and between x and $x+\frac{s}{2}$ on the other:

$$Hflucs(x,s) = |\overline{T_{[x-\frac{s}{2};\ x]}} - \overline{T_{[x;\ x+\frac{s}{2}]}}| \tag{5.1}$$

Finally, the Haar fluctuation for a given timescale s is defined as the root mean square of the fluctuations at all possible locations x on the timeseries. This result is then scaled by $\sqrt{\frac{\pi}{2}}$ which correspond to the mean of a half Gaussian distribution which $Hflucs$ follows (Lovejoy and Schertzer, 2012):

$$Hfluc(s) = \sqrt{\frac{2}{\pi} \times \overline{Hflucs(x,s)^2}} \tag{5.2}$$

Here, the data points of the timeseries are evenly spaced[2]. The fluctuations $Hflucs(x,s)$ are computed at the locations x corresponding to each data point present between $\frac{s}{2}$ and $N\Delta t - \frac{s}{2}$, where N is the number of data points and Δt the timestep of the timeseries. However, doing so results in overlapping fluctuations when s is two times above the timestep Δt. That is, a single data point can be used to compute several fluctuations $Hflucs(x,s)$ for a given scale s. While this helps reducing the uncertainty, it also under-represents the data points at each end of the timeseries. Finally, the Haar fluctuation $Hfluc(s)$ is computed for all scale s possible, that is, all even numbers between 2 and N. When the variable investigated is Gaussian (precipitation are Gaussian on monthly and longer timescales; Glasbey and Nevison, 1997), it is possible to convert the Haar fluctuations so they become proportional to

[2]Haar fluctuations can also deal with timeseries whose data points are not separated by a constant timestep, which is generally the case for proxy data.

a spectrum and can be compared to other spectrum estimates (Lovejoy and Schertzer, 2012), such as the multitaper method (Thomson, 1982, e.g. Figure 5.8-C):

$$spec(s) = s \times Hfluc(s)^2 \tag{5.3}$$

This method generally offers a good estimate of the spectrum's slope (Hébert et al., in press). Moreover, using the root mean square of the fluctuations instead of the standard deviation does not underestimate the spectrum of the longest scales as most other estimates of spectrum does.

Since overlapping fluctuations are considered, an analytical description of the uncertainty is not evident. Instead, the uncertainty is estimated from the sampling of 1000 white noise timeseries of the same length as the timeseries investigated. The Haar fluctuations are computed for each timeseries and the quantiles 0.05 and 0.95 define the 90% confidence envelope for each scale. The same fraction between the upper (lower) bound and the mean value is applied on the Haar fluctuations computed for the timeseries investigated.

5.4.1.2 Comparing the datasets

Figure 5.8 presents the Haar fluctuations (A) and the spectrum estimates (B and C) of the precipitation over the UIB, as an average over the winter period from October to May, and for the various datasets investigated. The Haar fluctuations (Figure 5.8-A) estimate how much the precipitation varies at different timescales. For example on the inter-annual timescale, ERA5 precipitation varies up to 125 mm. This value represents the average difference between two consecutive points, and therefore not related to the standard deviation (i.e. the average deviation from the mean, 94 mm here). It can be compared to the average precipitation of 504 mm. This value quickly decreases for larger timescales.

The power spectrum (B and C) allows a better understanding of the processes driving the variability. First, it can be noted that the Haar spectrum gives smoother results than the multitaper method, despite a log-smoothing. It highlights the lower noise level of the Haar fluctuations (see also the smaller confidence interval) which makes interpretations easier. ERA5 is the only dataset whose power spectra increases with the timescale, and only for the Haar spectrum (B), and is due to a negative trend in precipitation not captured by the multitaper method. However, the uncertainty is high, due to the small size of the timeseries, and GPCC, the other observational dataset, does not present the same behaviour. Beside uncertainty, this incease in power with timescales can also be related to the changing number and type of observations assimilated or the recent increase in the pace of global warming.

GPCC also presents some oscillations with peak variability in the sub-decadal and 20-30 years timescale, but within the confidence envelope.

Beside the transient simulation, the spectra are mostly flat, indicating a white noise. The difference in variability level can be mostly explained by the difference in mean precipitation. Surprisingly, the transient simulation presents a variability at the same level as the other simulations, despite drier conditions. It is possible that the dynamic vegetation adds more variability to the climate system, but other differences in the models can also explain that variation. Furthermore, the transient simulation shows an increase of the variability at timescale above 500 years, which is characteristic of the decrease in precipitation implied by the changing orbital forcing. This behaviour is missed when using multitaper (C) due to both the larger noise and the underestimation of the spectrum at the largest timescales.

The analysis of the differences between model data and observations is limited by observation uncertainties, yet inter-annual variability is clearly overestimated by the simulations. The drivers of such variability are investigated in the following section.

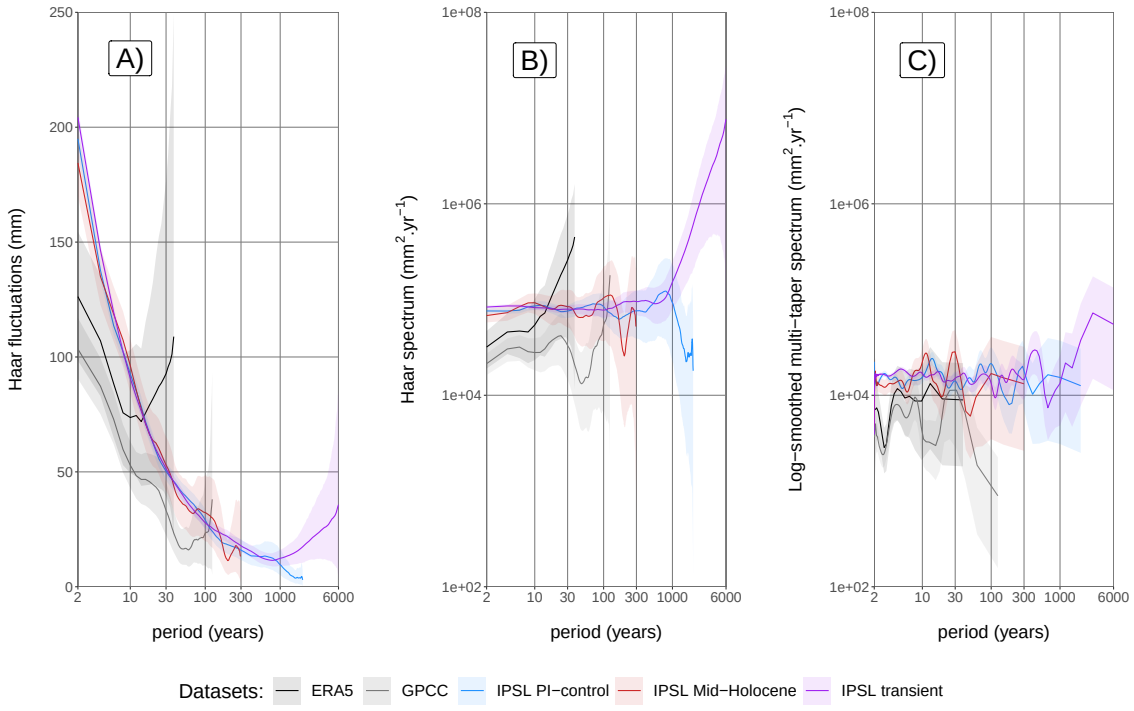


Fig. 5.8 Changes of variability across timescales for two observational datasets (ERA5 and GPCC) and three simulations (piControl, mid-Holocene, and transient): Haar fluctuations in A, spectrum estimates based on Haar fluctuation in B, spectrum estimates based on multitaper with the addition of a log-smoothing in C. The envelope indicates the 90% confidence interval.

5.4.2 Teleconnections

5.4.2.1 The importance of ENSO and NAO in the observations

The inter-annual variability of winter precipitation is dominated by ENSO and NAO (e.g. Filippi et al., 2014; Yadav et al., 2009), which depend on SST and SLP respectively. These two variables characterise the variability in the ocean and the atmosphere respectively but also the coupling between the two. Therefore, correlating the winter precipitation with SST and SLP gives a global view on the teleconnection patterns as well as on the sources of precipitation variability (e.g. atmospheric or oceanic sources).

Figure 5.9 shows the correlation maps between UIB precipitation and SST. October to May precipitation is correlated with the SST in December (left column), in order to compare it with the peak ENSO activity and in March (right column) to see the evolution of the teleconnection later in the season. The SST is provided by the observation-based HadISST dataset, while the precipitation data are from the two observational datasets (ERA5, A and B; and GPCC, C and D) using the maximal available period. The highest and positive correlations are found in the central equatorial Pacific, which reveals the influence of ENSO. The black box refers to the Niño3.4 box over which the ENSO intensity is traditionally computed (Trenberth, 1997). However, correlations are noticeably higher to the west of the box rather than to the east, which may suggest a stronger influence of Central Pacific rather than East Pacific El Niño events (cf. Trenberth and Stepaniak, 2001), which is the opposite to what Yadav et al. (2013) suggested for winter precipitation for northern India. Interestingly, no significant negative correlations are present around the tropical warm pool or to the south-west Pacific despite negative SST anomalies being traditionally found during El Niño events (Varotsos et al., 2013). By contrast, significant negative correlations are found in the northern Pacific. These correlations extend and get higher from December to March, which may reveal that the atmospheric circulation causing the more intense precipitation also impacts the SST field (e.g. change in the jet and in the mid-latitude disturbances in the Pacific). Similarly, an area of positive correlation develops in the equatorial Indian Ocean from December to March. The pattern is similar to a positive IOD (Indian Ocean Dipole) event in December (Saji et al., 1999), but does not resemble it in March, as IOD events generally decay during winter (Kug et al., 2006). Instead, this pattern in the Indian Ocean may well be the result of the ENSO forcing, as discussed in other studies (e.g. Klein et al., 1999). Interestingly, this area of positive correlation persists into the summer and moves north, which suggest an impact on the South Asia summer monsoon (not shown).

The correlation pattern between the two observational datasets ERA5 and GPCC, using their maximum time coverage, are very similar, but weaker in case of GPCC (maximum

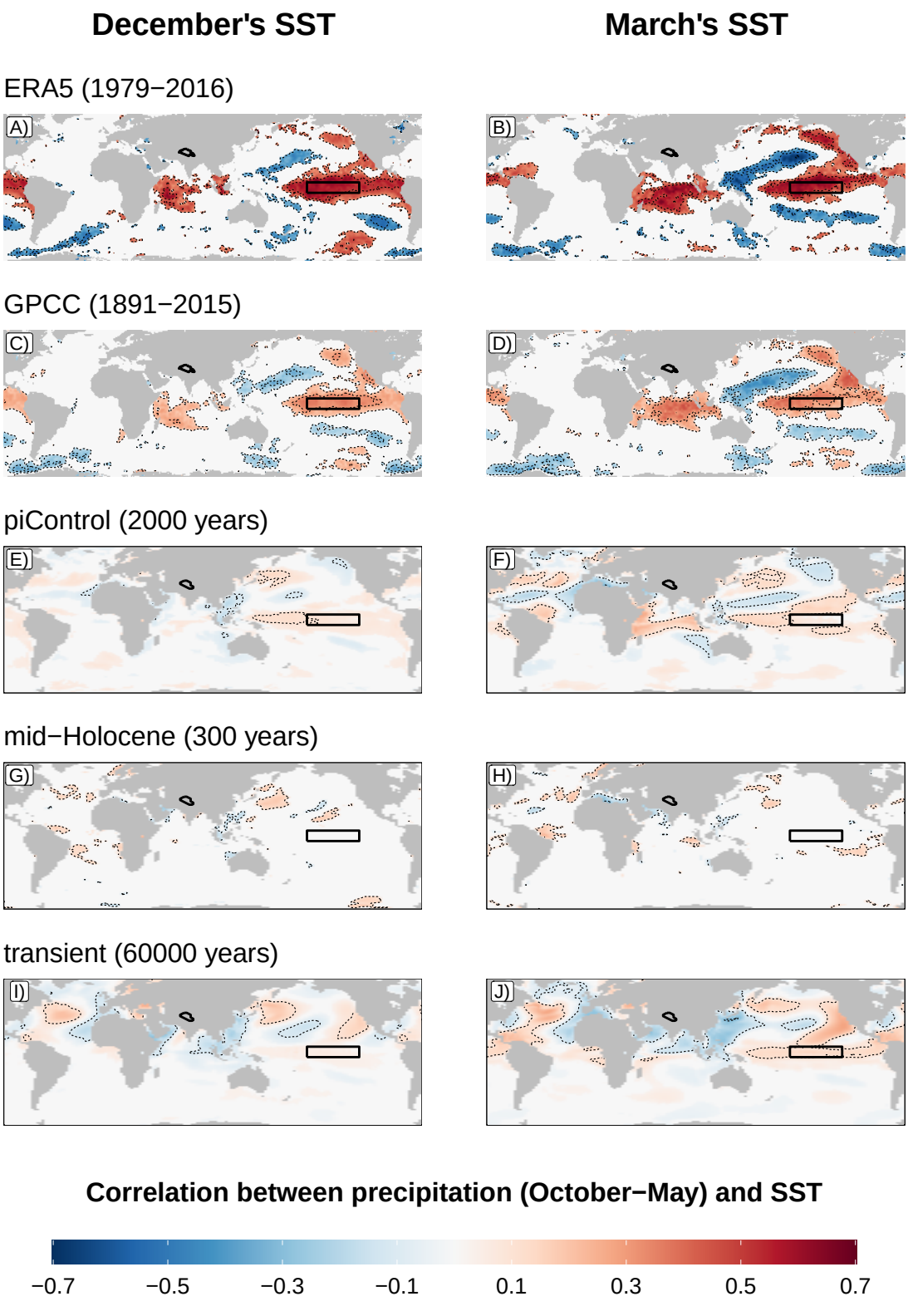

December's SST
March's SST
ERA5 (1979−2016)
A)
B)
GPCC (1891−2015)
C)
D)
piControl (2000 years)
E)
F)
mid−Holocene (300 years)
G)
H)
transient (60000 years)
I)
J)
Correlation between precipitation (October−May) and SST
−0.7
−0.5
−0.3
−0.1
0.1
0.3
0.5
0.7

Fig. 5.9 (Previous page) Correlation maps between winter (October-May) precipitation in the UIB (as indicated by the black contour line over land) and sea surface temperature in December A, C, E, G, I) and March (B, D, F, H, J) for the various datasets. Not significant values at the level 0.05 are shown in white. The black box in the middle of the Pacific is the Niño3.4 box used to compute the ENSO signal. The dotted lines are iso-contours as indicated by the labels in the legend. The timeseries are detrended before computing the correlation using a linear fit.

---

around 0.4) than ERA5 (almost 0.7). The longer time period in GPCC allows for a smaller threshold of the significant values which leads to pattern of similar extent as for ERA5. This difference is mostly explained by the change in the time period considered.

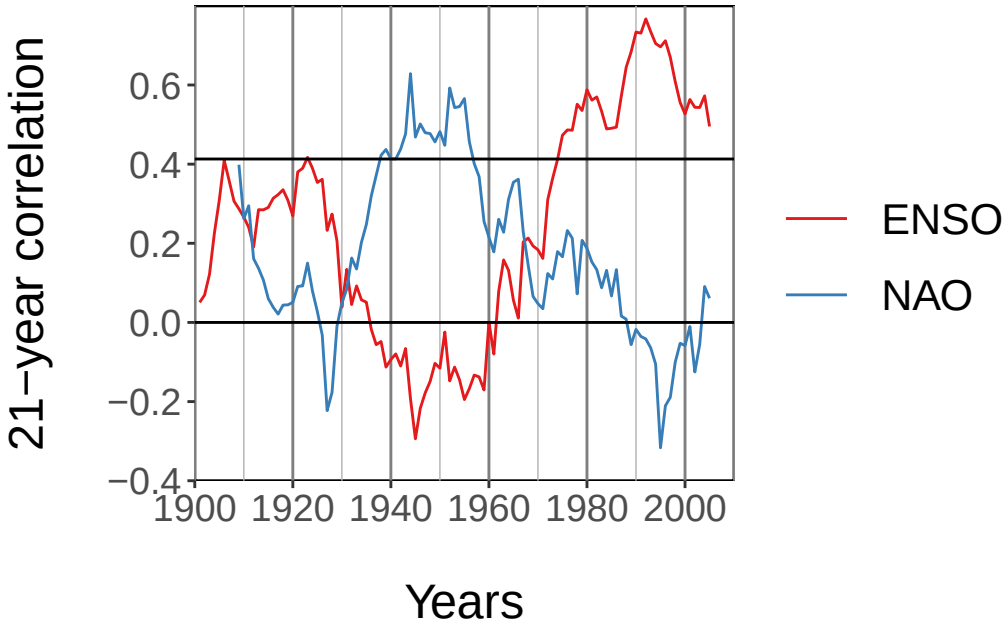


Fig. 5.10 Correlation between winter (October-May) precipitation in the UIB and the oceanic Niño index (ENSO, red) and the North Atlantic Oscillation index (NAO, blue), using a 21-year sliding window. The year corresponds to the centre of the window. Precipitation is based on GPCC, the ENSO index is computed from HadISST, and the NAO index is from observed pressure (PC-based; Hurrell et al., 2013). The threshold for significant positive values at the level 0.05 is 0.431 and is indicated by an horizontal line. The timeseries are detrended before computing the correlation using a linear fit.

Figure 5.10 shows the evolution of the correlation over 21 year long periods between the winter precipitation in GPCC and the ENSO signal, as measured by the 3-month average (NDJ) SST anomaly in the Niño3.4 box (ONI; Trenberth, 1997). It appears that ENSO only significantly (at the level 0.05) correlates with the precipitation starting from the 1970s, while the correlation sign is opposite, though not significant between 1930 and 1960. Correlations were higher but not significant before 1930. However, both GPCC and HadISST have a much higher uncertainty for this earlier period due to the limited availability and lower quality of observations, which likely reduces the correlations. This change in ENSO impact has been investigated before (Yadav et al., 2009).

Figure 5.10 also shows the evolution of the correlation between precipitation and NAO. The NAO index used here is based on the first principal component of observed North-Atlantic SLP (Hurrell, 1995; Hurrell et al., 2013). The October to March averages are considered, as for precipitation. Similar to ENSO, NAO exhibits multi-decadal variability, but with significant correlations only during the period between 1940 and 1960. Correlations are higher, and the significant period extends up to 1970, when considering averages over the shorter period from December to March for both precipitation and NAO, which then reproduces results from the literature (Filippi et al., 2014; Yadav et al., 2009). Interestingly, impact of ENSO and NAO on precipitation seems to alternate, so that NAO is not significantly correlated with precipitation after 1980. Data for a longer time period is needed to corroborate this result. Filippi et al. (2014) showed that the multi-decadal variation of NAO impacts could be related to a change in the longitudinal position of the centre of action (Icelandic low and Azores high) thus changing the Atlantic jet position and its ability to produce wave breaking as hypothesized in Section 4.5.2.

In Figure 5.11, the correlation between precipitation and SLP in ERA5 is investigated and shows no resemblance with NAO in the North Atlantic, in agreement with the result from Figure 5.10. Instead, the main pattern is a dipole of pressure between the tropical warm pool and the Eastern Pacific, which is typical of ENSO (i.e. displacement of the Walker circulation characterised by the southern oscillation). Most interestingly is the counter-intuitive extension of the very high correlation towards India and the Indus River basin: higher UIB precipitation are related to positive pressure anomalies in the Indus Plain, despite the supposedly stronger WDs. Indeed, the position of the pressure anomaly suggests weaker trade winds, or stronger south-westerly winds toward the UIB which helps to supply moisture to the UIB as shown in Syed et al. (2010) and Mariotti (2007) in the case of an El Niño event. This analysis is further supported by the higher correlation between the ENSO index and $Q700$ than with $W700$ contributing to precipitation as shown in the teleconnection summary in Table 5.2.

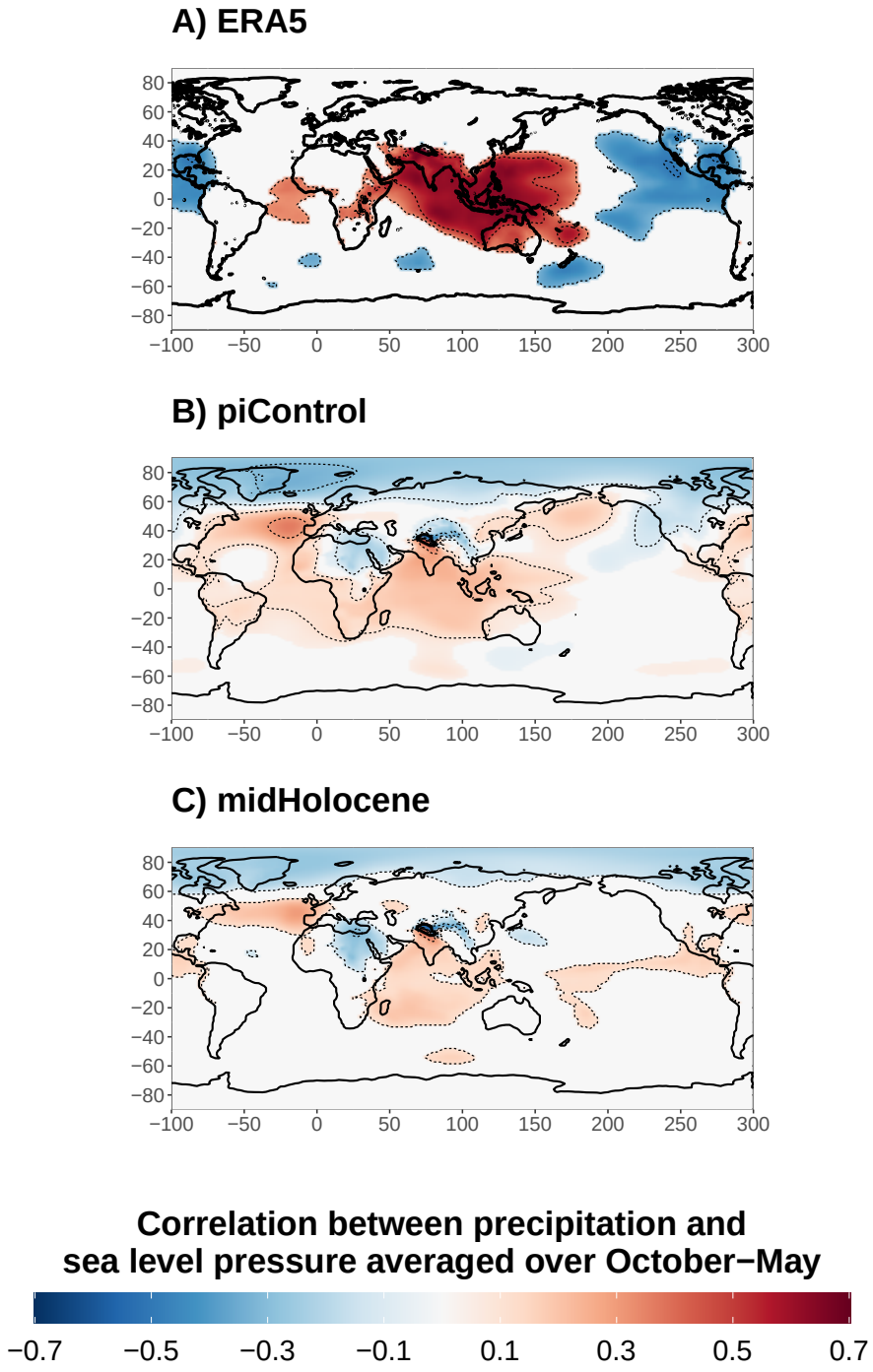


Fig. 5.11 Correlation maps between winter precipitation in the UIB and sea level pressure. Both values are averaged over the period October to May. Non-significant values at the level 0.05 are shown in white. The dotted lines are iso-contours as indicated by the labels in the legend. The timeseries are detrended before computing the correlation using a linear fit.

Table 5.2 Detrended correlations. Precipitation, positive $W700$ and $Q700$ are averages over the winter period (October to May). The ENSO timeseries is computed as the 3-month average (November to January) of SST anomaly in the Niño3.4 box, derived from HadISST for the observational datasets, and from modelled SST for the simulations. NAO is from Hurrell et al. (2013) for the observational dataset, and is computed from the first principal component of North Atlantic SLP for the simulations. In both cases NAO is an average over the winter period. Note that NAO is not available for the transient simulation. Non-significant values at the level 0.05 are in italic, significant ones in bold.

Dataset (Time period)	Variable	ENSO (NDJ)	NAO
ERA5 (1979-2017)	precipitation	**0.55**	*0.08*
	positive $W700$	*0.30*	*0.04*
	$Q700$	**0.59**	*0.06*
GPCC (1899-2015)	precipitation	**0.32**	**0.22**
piControl (2000 years)	precipitation	**0.09**	**0.28**
	positive $W700$	*0.04*	**0.29**
	$Q700$	**0.13**	**0.16**
midHolocene (6000 years)	precipitation	*-0.08*	**0.20**
transient (6000 years)	precipitation	**0.05**	

5.4.2.2 Correlations with SST and SLP in the simulations

Figures 5.9 and 5.11 also show the correlation between precipitation and respectively SST and SLP in the IPSL simulations.

The correlation with the ENSO index is below 0.1 for the piControl simulation (Table 5.2). Yet, the correlation is still significant due to the length of the timeseries (cf. Table 5.1). This low value is partly explained by the displacement of the bulk of highest correlations to the west of the Niño3.4 box (Figure 5.9-E and F). This westward displacement of the ENSO anomaly is typical for coupled global climate models and is due to mean SST biases (Bellenger et al., 2013), which have not been corrected in CMIP6 (Brown et al., 2020). It is possible that using an index based on the first principal component of equatorial Pacific SST to characterise ENSO would result in a better fit. The SLP correlation map confirms the small values and the westward displacement of the positive correlations: they extend over South Asia, the Indian Ocean and the tropical warm pool. The effect on the trade winds discussed in the previous section is further amplified by an area of negative correlation over Egypt and Arabia. In fact, ENSO is only correlated with $Q700$ and not $W700$ (Table 5.2), suggesting that, in the simulation, ENSO impacts moisture flux to the UIB, but not WD intensity.

By contrast, the correlation of precipitation with NAO is much higher than with ENSO (Table 5.2). This link is particularly evident on the SLP correlation map (Figure 5.11-B). On

this map, the NAO pattern, characterised by a dipole over the Atlantic, is the strongest feature and connects to the Arctic oscillation (dipole between the Arctic and the mid-latitudes). Furthermore, a tripole pattern is also evident in the North Atlantic on the SST correlation maps, particularly in March (Figure 5.9-E and F). The pattern is quite similar to the observed correlation pattern between SST and NAO (Visbeck et al., 2001). Also contary to the results for ENSO, NAO has higher correlation with $W700$ than with $Q700$, suggesting that, in the simulation, NAO impacts more WD intensity than the moisture field (Table 5.2).

Finally, correlations are overall higher in March than in December, which suggests that the simulated SST rather reacts to the atmospheric forcing, such as NAO, than acts on it, at least in the context of winter precipitation in the UIB.

Comparing the strength of the teleconnection in the piControl simulation and in the observations is complex as the observations exhibit important multi-decadal variability and are not available over a long enough period of time. Yet, when comparing the piControl simulation with GPCC, the simulated winter UIB precipitation has stronger correlation with NAO but weaker correlation with ENSO (cf. Table 5.2). Furthermore, Figure 5.12 attempts to investigate multi-decadal variability in the piControl simulation: 21-year correlations between precipitation and ENSO or NAO seem to exhibit some decadal variability, but more work is needed to find out whether this variability is just statistical noise or a modulation of the teleconnections. Yet, significant and positive correlations of precipitation with ENSO and NAO respectively appear mutually exclusive, as in the observations.

Finally, the teleconnections in past simulations are investigated. In the mid-Holocene simulation, there is no evident pattern of correlation between precipitation and SST (Figure 5.9-G and H). This absence is quite suspicious, but may also be related to the shorter time period used to compute the correlations (300 years, instead of 2000 in the piControl). As for SLP, patterns are similar to those in piControl, but weaker (cf. also Table 5.2).

In the transient Holocene simulation, patterns of correlation between SST and precipitation are somewhat different to those of the piControl (Figure 5.9-I and J). The teleconnection with ENSO is even lower in December (cf. Table 5.2), but further increases in March. Different patterns are also evident in the North Atlantic, North Pacific and Indian Oceans. Further research is needed to find out whether these differences are due to the changes in the model or to the transient forcings.

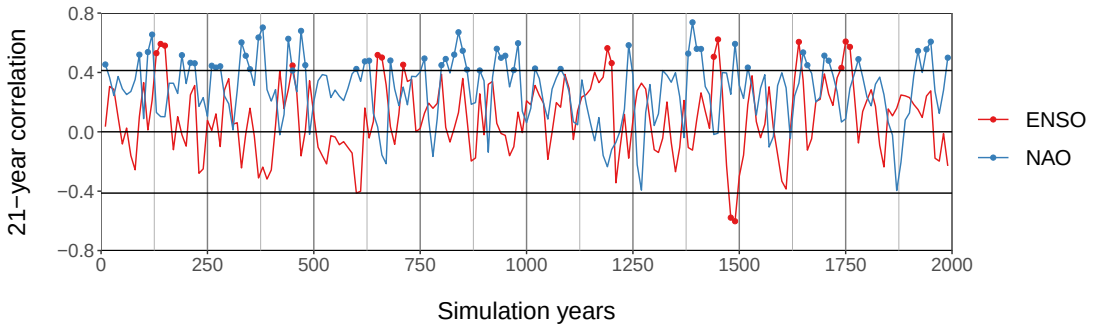


Fig. 5.12 Same as Figure 5.10 but based on data from the piControl simulation. Significant values at the level 0.05 are further highlighted by dots. Note that the correlations are computed every 10 years instead of every year as in Figure 5.10.

5.5 Conclusion

This chapter presents an evaluation of the simulated winter precipitation in the Upper Indus River basin (UIB) in the IPSL global climate model used for CMIP6 (IPSL-CM6A-LR). The model represents the driver of winter precipitation such as the WDs, but different biases have been noted: an overestimation of precipitation, a dryer atmospheric column and a stronger Subtropical Westerly Jet.

In winter the simulated jet maximum anomaly is located south of the Himalayas and further increases towards the end of the season, between March and April: biases in the heat source of the Tibetan Plateau are likely to play a role but more analysis is needed to test this hypothesis. Spatial resolution also may play a role in the transfer of kinetic energy from the polar jet to the subtropical jet. The stronger subtropical jet intensifies the WDs, which in turn increase cross-barrier winds in the UIB.

The dry bias in atmospheric humidity in the UIB is a regional feature but it is particularly pronounced at subtropical latitudes and related to a change in average moisture transport. Stronger trade winds transport more moisture to the southern hemisphere while the westerly transport is pushed northward. While this process may be related to the jet anomaly, it also suggests local changes in the Hadley circulation which are complex to investigate.

Finally, the precipitation bias is not related to a change in average wind or humidity: these biases almost cancel each other out. Instead, sub-daily data shows that the simulated wind field associated with a WD is different from the reanalysis: the meridional flow is stronger compared to the zonal flow. Particularly, the southerly wind ahead of a WD originates from further south, and is therefore able to produce a stronger moisture anomaly. Smoother representation of the relief and possibly the bias in the subtropical jet may explain this behaviour. This analysis highlights the importance of the availability of 3D data at sub-daily resolution for at least a part of the simulation.

These biases leads to some differences between the simulated seasonality of winter precipitation and the observations. The main difference is a delayed precipitation peak while WDs remain active until June, due to the stronger jet stream in the late winter season. This delay, combined with the high positive precipitation bias in autumn (October to December), limits the possible extent of the second precipitation peak during summer. Hence, this analysis of the winter precipitation biases may offer new avenues to understand and correct the greater issue of most climate models: the representation of the summer monsoon.

The simulated teleconnections are challenging to evaluate since observations show a multi-decadal variability in the teleconnections strength. It seems the IPSL-CM6-LR model overestimates the atmospheric teleconnections such as the link with NAO, but underestimates

the correlation with SST and particularly ENSO. It is unclear whether the model is able to reproduce multi-decadal variability in the teleconnections strength. Furthermore, the simulation does not represent specific modes of variability at the interdecadal to multidecadal scale. This limitation has been discussed in the literature (Mann et al., 2020) and can be seen as a result of the weak coupling with the surface ocean. No modes of variability are evident in observations either, although uncertainty is high: climate archives are needed to investigate these longer timescales.

The bias in seasonality is particularly problematic when investigating past and transient simulations, such as for the Holocene. The change in seasonality due to external forcing is small compared to the biases. Furthermore, if the too weak coupling with the ocean is confirmed, the model may also underestimate the seasonality change due to changes in ocean circulation. Yet, interesting trends have been found: less precipitation during the mid-Holocene during the core winter season (January-February) but more precipitation in the early and later part of the WD season (October-November and April-May).

The study presented here highlights the general challenges of climate modelling and offers the basis for a multi-model evaluation of precipitation in the UIB, and potentially for other parameters elsewhere.

Chapter 6

Summary and conclusion

6.1 Thesis aim and approach

This thesis has aspired to provide archaeologists with climate information so that they can further their understanding on the development and demise of the first urban society in the Indus River Basin, which flourished between ca. 4600 and 3900 BP (Sections 1.1 and 1.3). Rainfall is the most impactful variable to the surface environment and the human societies of the plains of the Indus River Basin (Section 1.3.2). In addition, a dense river network irrigates the northern plain, thanks to direct rainfall and snowmelt run-off from the mountains (Sections 1.2.2 and 1.2.3). Thus, precipitation in the whole Indus River Basin has been considered. Furthermore, the various timescales of precipitation are associated with different human activities and population behaviour (Section 1.5 and Figure 1.8). Hence, the thesis has reconstructed precipitation variability (Chapters 3, 4 and 5), and used this knowledge to discuss the precipitation variability across timescales during the middle to late Holocene in the Indus River Basin (Chapter 5).

Traditionally, information about past climate is derived from climate archives (cf. Section 1.4.1). However, this method has disadvantages in this context, particularly due to the absence of archives near archaeological sites, or the difficulty to distinguish the variability of the two wet seasons of the Indus River Basin (Section 1.1). With these limitations in mind, the thesis has focused on the potential added-value of climate model simulations to climate archives such as the denser spatial coverage and the resolution of the seasons. Consequently, it has revolved around the evaluation of global climate model outputs (Chapters 3 and 5), and the methods to perform it (Chapters 2 to 5).

6.2 Thesis structure

The first aspect of the thesis (Chapter 2) has concerned the choice of a reference dataset among the observational datasets (including reanalyses), used for the rest of the study (Chapter 3 to 5). It has distinguished the northern Upper Indus Basin (UIB) for the southern lower part (LIB). It has involved a cross-validation of the precipitation datasets and a quantification of the present day amount of precipitation falling in the Indus basin, and its uncertainty.

The second aspect (Chapters 3, 4 and 5) has explored precipitation variability in the observations at various time and spatial scales: from a 3-hourly and local (Chapters 3 and 4) to an inter-annual and global scale (Chapter 5). This analysis has served two goals. First it has provided information on precipitation variability at these scales, and second, it has designed methods to test the drivers of such variability in climate models. This analysis has been performed under the assumption that the change of precipitation across the Holocene is due to a modulation of that higher frequency variability rather than a more dramatic change in drivers. Arguably, the changes in boundary conditions during the Holocene (e.g. orbital forcing, GHG concentration, sea and land surface conditions) are not large enough to significantly transform oceanic and atmospheric circulations (Section 1.4.2.1; Wanner et al., 2008).

This second aspect has discussed the variability at three different spatial and temporal scales. First, it has analysed what triggers the 3-hourly precipitation at the scale of the UIB (Chapter 3). Second, it has described the atmospheric circulation in the days preceding and following a precipitation event in winter in the UIB (Chapter 4). Finally, it has addressed the question of global and inter-annual teleconnections in relation to winter UIB precipitation (Chapter 5).

The third aspect of the thesis has been an analysis of climate model outputs. The evaluation of global climate models (Chapter 3 and 5) has been performed using the knowledge on present variability developed in the second part of the thesis. Then, past climate simulations have been critically investigated based on the result of the present day evaluation (Chapter 5).

6.3 Methods

The thesis has been based on statistical tools and methods designed to investigate timeseries. Some of these methods are not regularly used in atmospheric science and are highlighted here.

In Chapter 2, the variability of observational precipitation datasets is validated using a correlation analysis (Sections 2.3.3 and 2.3.4). Usually, correlations are computed between a

trusted reference dataset and the datasets being investigated, thus the estimated quality of these datasets is biased towards agreement with the reference rather than reality. The novel approach presented here is to alternatively use all datasets as a reference instead. The method particularly makes use of the independence between most reanalyses and other observational datasets (e.g. rain gauge datasets). Indeed, more accurate reanalyses have a variability closer to rain gauge datasets. Similarly, more accurate rain gauge datasets have a variability closer to reanalyses. This approach can be used to order the different datasets depending on the quality of their representation of the precipitation variability.

In Chapters 3, 4 and 5, the importance of moisture transport to explain precipitation variability is quantified using a regression analysis (cf. Sections 3.3 and 4.3). This analysis is usually performed by computing the regression coefficients (or the correlation coefficients) between the precipitation timeseries investigated, and the moisture transport at each grid point (similar to Figure 5.1). This method indicates the importance of moisture transport at each grid point, and shows the pattern of moisture transport associated with an increase of precipitation. However, it does not quantify the importance of the moisture transport field as a whole, and most of the significant regression coefficients can be explained by the spatial auto-correlation of the moisture transport field, rather than by a direct link with the precipitation. These issues are solved by using a multi-linear regression, and more specifically a PC regression (Section 3.3.2). The method is further explored to investigate the use of several predictor fields, and how to select the most important ones (Section 3.3.3). The PC regression is eventually optimised for precipitation characteristics (i.e. precipitation are always positive, Section 4.3). This method is not specific to precipitation and moisture transport and can in fact be used in a wide range of contexts (e.g. Section 4.4.2).

The dependence between variability and longer timescales is investigated in Section 5.4.1. Spectral density are typically estimated using a periodogram, or more complex methods such as multitaper, which present limitations (Section 5.4.1.1). A more intuitive method is used instead, the Haar fluctuation (Section 5.4.1.1). Work on the Haar fluctuations to investigate variability is recent, but present promising results compared to other estimates of spectral density (Section 5.4.1.2).

6.4 Results, limitations and future avenues.

6.4.1 Observational precipitation datasets

Investigating a large variety of observational datasets shows that large uncertainties exist on the precipitation amount and pattern in the Indus River Basin, particularly in the mountainous

part (Sections 2.3.1). Considering area-wide average (UIB, LIB) limits the impact of local scale discrepancies but biases remain. In comparison to a recent study using a high density of rain gauges from the mountains (Dahri et al., 2018), all gridded rain-gauge based datasets underestimate precipitation, mostly due to a biased location of measurements and wind-driven snowfall underestimation (2.3.1.3). In fact, only GPCC takes into account wind-driven snowfall underestimation and probably not sufficiently. For example, the common use of APHRODITE as a reference is problematic as it appears to be one of the driest datasets for the whole Indus River Basin. Future interpolation methods used to produce gridded datasets from station measurements need to be updated to consider higher precipitation at altitude and to account for measurements biases.

Reanalyses offer a higher estimate of the precipitation than rain gauge-based datasets, possibly closer to the reality although overestimations cannot be ruled out (Section 2.3.1.2). They agree well on the amount of winter UIB precipitation when considering area-wide average but summer precipitation have a larger range (Section 2.3.2). Similarly, the confidence on winter precipitation variability is high, as good as other observational datasets, but is more dependent on the dataset in summer (Section 2.3.3.3). Twentieth century-long reanalyses perform worse than other reanalysis, particularly in winter, and do not reproduce any of the decadal variability showed by the observations (Section 2.3.4). By contrast ERA5 is the only reanalysis that matches the observations variability on the period 1979-2018 from daily to decadal variability for both winter and summer (Section 2.3.4) and is therefore suited for the analyses in the thesis. Release of ERA5 for the period 1950-1978 in September 2020 offers new opportunities for longer analyses if the dataset retains the same quality for the whole period.

This analysis highlights the need to refer to or perform a cross-validation of observational datasets before using any, and particularly when interpreting the results of climate models. It also highlights the difficulty to validate local scale patterns. Hence, the thesis has focused on basin scale precipitation. It has made use of ERA5 (1979-2018) to investigate present day variability (Chapter 3, 4, 5), completed by GPCC for the inter-annual timescale due to its longer time period (Chapter 5).

6.4.2 Drivers of precipitation variability

6.4.2.1 Cross-barrier moisture transport

The local driver of 3-hourly precipitation variability in the UIB was studied using a PC regression with moisture transport (Section 3.4.1). The regression showed that more than 80% of the precipitation variability is related to southerly moisture transport at 850 hPa and

700 hPa along the Himalayan foothills. Hence, precipitation is mostly triggered by the forced uplift of a cross-barrier flow. This process explains both wet seasons, and the decomposition of the precipitation signal exhibits seasonal differences (Section 3.4.2). The winter wet season is dominated by moisture transport at higher altitude, and is triggered by an increase in winds. By contrast, the summer wet season is related to an increase in moisture at both altitudes, while the wind increase is of smaller importance. The method can be reproduced for climate models to test precipitation triggers as well as to investigate biases in the seasonality (Section 3.4.5).

A similar technique can be used for the southern part of the Indus River Basin, to determine for example the altitude and direction of moisture transport that is the most important to trigger convergence. However, the method is not very well suited for convective processes which would need to be investigated at lower spatial scale. Furthermore, the importance of latent heat released to sustain the vertical uplift has not been quantified and may also be a subject for further analysis. Finally, static stability and frontal activities may play a role in the precipitation pattern and the balance between precipitation on the plain or in the mountains. This balance has a significant impact on human activities as rain in the plains directly feeds crops while precipitation in the mountains flows into the rivers and ephemeral channels, and additional irrigation works may have been needed to develop cropping activities using that water source.

6.4.2.2 Western Disturbances

Winter precipitation falls in the UIB during events between October and May (Section 4.7.1) lasting between several hours to a few days and are triggered by synoptic atmospheric disturbances called Western Disturbances (WDs, Chapter 4). WDs are primarily cold core upper-level (300 hPa) cyclonic disturbances that extend towards the ground with a tilt characteristic of baroclinic processes (Section 4.4.1). Precipitation rates are dependent on various characteristics of WDs. The 700 hPa cross-barrier winds responsible of the precipitation in winter are triggered by the approach of a WD from the west of the UIB. Hence, these winds are dependent on the intensity of the WDs and more specifically on the zonal gradient of geopotential at 300hPa across the UIB (i.e. southerly geostrophic wind at 300 hPa, Section 4.1). Equally important is the thermal structure of the WD, which constrains the extension of southerly winds from 300 hPa to 700 hPa (Section 4.4.3). With its cold core, a WD also drives a zonal temperature gradient across the UIB, which inhibits lower level winds. However, this gradient can be reduced by stronger baroclinicity, mixing due to the uplift and latent heat release. Finally, WDs also drive the moisture anomaly at 700 hPa through the UIB cross-barrier winds, the larger atmospheric circulation, and the WD

history (Section 4.6). The most southern WDs, from the Arabian Peninsula, associated with stronger southerly winds along their track, lead to the highest positive anomaly in moisture. By contrast, WDs from the north-west, associated with stronger northerly rear advection drive drier air to the Indus River Basin, in particular directly after the peak in cross-barrier winds.

The position and intensity of the Subtropical Westerly Jet (SWJ, 250hPa) impact the characterictics of WDs and can explain the wind seasonality (Section 4.7.3). A more rapid SWJ fuels deeper WDs, hence the SWJ peak intensity in January explains the WD peak intensity in the same month. A more conducive thermal structure (and therefore stronger 700 hPa cross-barrier winds) develops when the SWJ maximum intensity occurs to the south-west of the WD. In the opposite case, the maximum intensity occurs to the south-east of the WD. The SWJ is stronger to the south-east of the UIB between October and January and forms an extension of the East Asian Jet, while, between February and May, a second jet maximum develops to the west of the UIB, the Arabian jet. Hence, for a similar WD intensity, the cross-barrier winds are reduced in autumn, but increased in spring.

Finally, the seasonality of moisture at 700 hPa in the UIB can be explained by the total column moisture transport (Section 4.7.4). The minimum average moisture occurs in late December, when trade winds in the Arabian Sea and western Indian Ocean are the strongest, sending surface moisture to the southern hemisphere. The ensuing weakening of the trade winds leads to the establishment of a westerly moisture pathway from the Red Sea, across the Persian Gulf and into the north of the Arabian Sea. The passing of WDs reinforces this pathway and steers it towards the UIB.

This comprehensive description of the synoptic variability of winter UIB precipitation presents new explanations for precipitation seasonality and therefore new ways to investigate seasonal biases in climate models, through the representation of the WDs. Potential avenues include the impact of upper-level disturbances during the summer months, as well as the origin of the synoptic variability of summer precipitation in both the southern plains (LIB) and the northern slopes (UIB) of the Indus basin. The archaeological community would also benefit from a better understanding on the finer spatial scale precipitation variability associated with WDs (e.g. plain-mountain balance, Section 6.4.2.1).

6.4.2.3 Teleconnections

Winter UIB precipitation exhibits inter-annual variability triggered by primarily atmospheric and extra-tropical teleconnections such as the North Atlantic Oscillation (NAO), or coupled oceanic-atmospheric and tropical teleconnections such as the El Niño/Southern Oscillation (ENSO, Section 5.4.2.1). The NAO influence originates at the intra-seasonal frequency and

mostly affects the cross-barrier winds in the UIB through the development of stronger WDs, in relation to wave breaking and possibly blocking events (Section 4.5.2). By contrast, ENSO mostly affect the moisture transport from the Arabian Sea through a weakening of the trade winds. Further work is needed to understand the influence of NAO and blocking events, as well as to explain the impact of ENSO and its different flavours. Furthermore, the impact of these teleconnections on the seasonality and on the summer monsoon specifically needs to be investigated.

NAO and ENSO teleconnections exhibit a seemingly opposite multi-decadal variability since the start of the instrumental records, which impacts the lower frequency of the precipitation variability. The decadal to multi-decadal variability of the precipitation could be further explored using the extension of ERA5 data to 1950.

6.4.3 Global climate model evaluation

6.4.3.1 Present day variability

UIB precipitation is investigated in the historical simulations of three CMIP6 climate models (IPSL, MRI, GISS, Section 3.4.5). The three overestimate ERA5 precipitation between October and May but do not simulate the summer peak, a bias common to many CMIP5 simulations. The three simulations successfully represent the relationship between cross-barrier moisture transport and precipitation, but the seasonality of cross-barrier moisture transport is biased. The origin of the biases in winter is further explored in the Pre-Industrial-control (piControl) simulation of IPSL, which simulates a precipitation seasonality similar to the historical run. The piControl simulation underestimates moisture content at 700 hPa and overestimates 700 hPa cross-barrier winds, so that the two biases cancel each other out (Section 5.3.2). The positive bias in moisture transport is instead explained by a stronger spatial and temporal correlation between cross-barrier winds and the moisture anomaly. Particularly, the simulated WDs do not have exactly the same characteristics, with stronger meridional winds bringing more moisture from further south in the Arabian Sea, possibly because of the coarse resolution of the relief (Section 5.3.3). Furthermore, the simulated WDs are stronger, explaining the cross-barrier wind overestimation, and is related to a more rapid SWJ. The SWJ anomaly is the highest in spring, effectively extending the activity of the WDs up to June, and delaying the precipitation peak to March-April compared to ERA5. Such a behaviour and the spatial pattern of 250 hPa wind anomaly suggest weaker or delayed establishment of the spring heat sources from the Tibetan Plateau that needs further investigation. By contrast, the lower level westerly moisture pathway between the Red Sea

and the Arabian Sea is not simulated, while stronger trade winds transport more moisture to the southern hemisphere, resulting in the drier atmosphere discussed earlier.

The analysis of the teleconnections (Section 5.4.2.2) shows that the winter UIB precipitation in the IPSL piControl simulation is very weakly correlated with ENSO and SST patterns in general and that these correlations are higher towards the end of the season, suggesting that the SST patterns are mostly a response to the anomalous atmospheric circulation. By contrast, precipitation is more strongly correlated with NAO and SLP patterns in general. Comparison with observations is complex, due the multi-decadal variability of the teleconnection strength in the observations and the possible influence of the anthropogenic global warming. More analyses are needed to disentangle low frequency variability from statistical noise.

Spectrum analysis (Section 5.4.1.2) indicates that the IPSL simulation overestimates precipitation variability mostly due to the increased precipitation amount, and that no modes of variability are evident (that is, different from a white noise). No modes of variability are evident in the observations either, but they may be hidden due to large uncertainties. Moreover, the simulation is not affected by any changing external forcing (e.g. anthropogenic emissions, volcanic eruptions) and does not represent key components of the earth system such as vegetation or the carbon cycle which could add other sources of variability on larger timescales. To conclude, the IPSL piControl simulation exhibits various biases that transform UIB precipitation seasonality, while teleconnections and variability on larger timescales are difficult to validate. This evaluation helps to better understand the origin of the biases so these biases can be corrected in the future development of the model, although more analyses are needed regarding the summer monsoon. However, these results are specific to the IPSL-CM6A climate model and cannot be extrapolated to others. Instead, this analysis develops a framework to perform multi-model evaluations that was only partially conducted with GISS and MRI, and that could help to understand and characterise the biases that all climate models exhibits as shown by previous studies.

6.4.3.2 Past climate simulation

The Holocene precipitation was investigated using two IPSL simulations (Chapter 5), one using the orbital forcing corresponding to 6000 years BP (mid-Holocene), which can be compared to the piControl simulation. The other simulation models the last 6000 years using transient orbital forcing and greenhouse gas concentrations, and includes dynamic vegetation. As discussed in Sections 1.4.3, vegetation, soil cover and dust emission could have had a large impact on the monsoon precipitation in Africa and Asia during the mid-Holocene. However, the equilibrium mid-Holocene experiment uses the same conditions as the piControl for those

variables. By contrast, the transient simulation accounts for the vegetation and soil cover change by simulating it dynamically.

Both transient and mid-Holocene simulations suggest a similar multi-millenial change in precipitation seasonality, although they disagree on its intensity: the core winter season is drier compared to the present, but the summer monsoon, and the early and late part of the winter season wetter. The influence of the change in orbital forcing on the summer monsoon is well documented in both precipitation proxies and theory (cf. Section 1.4.1 and 1.4.2.1). However, its influence on the winter season in less clear and needs further investigation with a multi-model approach. For example, the drier core winter season may be related to cooler conditions which decrease moisture content. A previous study (Hunt and Turner, 2019) discussed a change in SWJ intensity that was not apparent in the mid-Holocene simulation investigated. Furthermore, these changes in precipitation seasonality are small compared to the biases which significantly decreases the confidence in the results (Section 5.3.4).

The simulations do not suggest a change in the teleconnection patterns or in the inter-annual to centennial variability (Section 5.4.2.2). However, these findings are difficult to validate. Particularly, a very weak coupling between the surface ocean and the precipitation is observed, which may impact the mean response to orbital forcing (i.e. climate sensitivity) and limit the precipitation variability change at inter-annual to multi-decadal timescale. By contrast, climate records suggest multi-decadal oscillations around the time of the end of the Indus civilisation (e.g. 4.2 ka, Section 1.4.1). Some of this variability may derive from the gradual changing of orbital forcings and therefore cannot be represented in simulations with fixed orbital parameters such as the piControl run or the mid-Holocene run. However, even the transient simulation with dynamic vegetation does not show any specific multi-decadal variability. Other external forcing such as the Solar activity and volcanic eruptions can impact multi-decadal variability but remain poorly understood (Section 1.4.2.1). A better representation of the present climate mean and variability as well as the regional multi-millenial Holocene trends (e.g. Sahara greening) in climate models is a pre-requisite to investigating multi-decadal variability resulting from orbital changes. Although these results are specific to a family of climate models, the limitations unravelled here are common to all climate simulations. Hence, at their present state, past-climate simulations do not provide the nuanced information on precipitation changes and variability needed in an archaeological context, and will not until significant breakthroughs are achieved. Those breakthroughs include a more realistic representation of the present day climate (e.g. parametrisation of subgrid processes, land-atmosphere and ocean-atmosphere coupling) and the understanding of the drivers of variability at the millennial to decadal scale.

6.5 Conclusion

Global Climate Models (GCMs) are tools designed by and for climatologists to investigate the processes affecting the global climate system. By contrast, archaeologists uses local climate and more generally environmental data as input to investigate societal systems and human behaviours. Hence, GCMs focus on providing interpretable climate data, which is key for any climate studies, while plausibility of the data and estimation of the uncertainty are most important in an archaeological context. Furthermore, at the spatial scale investigated by archaeologists, GCMs often produce unrealistic means and seasonality, and questionable variability and changes of the climate variables across the Holocene. In conclusion, GCMs are not particularly suited for archaeological purposes. Regional climate models help to investigate phenomena on a more local scale, such as surface change and convection, but still suffer from the lack of plausibility, if past climate simulations are used for boundary conditions. Instead, present day observations often offer better estimates of the past mean, seasonality and high frequency variability (e.g. days to decades), even considering potential changes across the Holocene.

An ideal dataset would combine the realisticness of present day observations, information on past climate from proxy data and the spatial and temporal constraints of physical laws used in climate models. Deep learning methods may be able to achieve this (Reichstein et al., 2019), and should be explored in the future.

References

Achyuthan, H., Kar, A., and Eastoe, C. (2007). Late Quaternary-Holocene lake-level changes in the eastern margin of the Thar Desert, India. *Journal of Paleolimnology*, 38(4):493–507.

Adam, J. C. and Lettenmaier, D. P. (2003). Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research: Atmospheres*, 108(D9).

Adler, R., Sapiano, M., Huffman, G., Bolvin, D., Wang, J., Gu, G., Nelkin, E., Xie, P., Chiu, L., Ferraro, R., Schneider, U., and Becker, A. (2016). New Global Precipitation Climatology Project monthly analysis product corrects satellite data shifts. *GEWEX News*, 26(4):7–9.

Agnihotri, C. and Singh, M. (1982). Satellite study of western disturbances. *Mausam*, 33(2):249–254.

Ahmed, F., Adnan, S., and Latif, M. (2019). Impact of jet stream and associated mechanisms on winter precipitation in Pakistan. *Meteorology and Atmospheric Physics*, 132(2):225–238.

Aizen, E. M., Aizen, V. B., Takeuchi, N., Mayewski, P. A., Grigholm, B., Joswiak, D. R., Nikitin, S. A., Fujita, K., Nakawo, M., Zapf, A., and Schwikowski, M. (2016). Abrupt and moderate climate changes in the mid-latitudes of Asia during the Holocene. *Journal of Glaciology*, 62(233):411–439.

Al, W., Orking, G., and Clima, O. (2008). Climate change and food security: a framework document. Technical report.

Ali, G., Rasul, G., Mahmood, T., Zaman, Q., and Cheema, S. (2012). Validation of APHRODITE precipitation data for humid and sub humid regions of Pakistan. *Pakistan Journal of Meteorology*, 9(17).

Andrews, J., Carolin, S., Peckover, E., Marca, A., Al-Omari, S., and Rowe, P. (2020). Holocene stable isotope record of insolation and rapid climate change in a stalagmite from the Zagros of Iran. *Quaternary Science Reviews*, 241:106433.

Antonopoulos, J. (1992). The great Minoan eruption of Thera volcano and the ensuing tsunami in the Greek Archipelago. *Natural Hazards*, 5(2):153–168.

Archer, D. R. and Fowler, H. J. (2004). Spatial and temporal variations in precipitation in the upper indus basin, global teleconnections and hydrological implications. *Hydrology and Earth System Sciences Discussions*, 8(1):47–61.

Arz, H. W., Lamy, F., and Pätzold, J. (2006). A pronounced dry event recorded around 4.2ka in brine sediments from the northern Red Sea. *Quaternary Research*, 66(3):432 – 441. Holocene Climate and Cultural Evolution in Late Prehistoric-Early Historic West Asia.

Ashok, K., Guan, Z., Saji, N. H., and Yamagata, T. (2004). Individual and combined influences of ENSO and the Indian Ocean Dipole on the Indian Summer Monsoon. *Journal of Climate*, 17(16):3141–3155.

Ashutosh, S. (2019). India state of forest report. Technical report.

Azharuddin, S., Govil, P., Singh, A., Mishra, R., Agrawal, S., Tiwari, A., and Kumar, K. (2017). Monsoon-influenced variations in productivity and lithogenic flux along offshore Saurashtra, NE Arabian Sea during the Holocene and Younger Dryas: A multi-proxy approach. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 483:136 – 146. Development, evaluation and application of marine paleoclimatic/paleoceanographic proxies: An update.

Bao, J.-W., Michelson, S. A., Neiman, P. J., Ralph, F. M., and Wilczak, J. M. (2006). Interpretation of Enhanced Integrated Water Vapor Bands Associated with Extratropical Cyclones: Their Formation and Connection to Tropical Moisture. *Monthly Weather Review*, 134(4):1063–1080.

Barriopedro, D., García-Herrera, R., Lupo, A. R., and Hernández, E. (2006). A climatology of northern hemisphere blocking. *Journal of Climate*, 19(6):1042–1063.

Barros, A., Kim, G., Williams, E., and Nesbitt, S. (2004). Probing orographic controls in the himalayas during the monsoon using satellite imagery. *Natural Hazards and Earth System Sciences*, 4(1):29–51.

Baudouin, J.-P., Herzog, M., and Petrie, C. A. (2020a). Contribution of Cross-Barrier Moisture Transport to Precipitation in the Upper Indus River Basin. *Monthly Weather Review*, 148(7):2801–2818.

Baudouin, J.-P., Herzog, M., and Petrie, C. A. (2020b). Cross-validating precipitation datasets in the Indus River basin. *Hydrology and Earth System Sciences*, 24(1):427–450.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I., Huffman, G. J., Adler, R. F., and Wood, E. F. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, 23(1):207–224.

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., and Vialard, J. (2013). ENSO representation in climate models: from CMIP3 to CMIP5. *Climate Dynamics*, 42(7-8):1999–2018.

Berger, A. (1988). Milankovitch theory and climate. *Reviews of Geophysics*, 26(4):624–657.

Berger, A. and Loutre, M. (1991). Insolation values for the climate of the last 10 million years. *Quaternary Science Reviews*, 10(4):297 – 317.

Berkelhammer, M., Sinha, A., Stott, L., Cheng, H., Pausata, F., and Yoshimura, K. (2013). An abrupt shift in the Indian monsoon 4000 years ago. In *Climates, Landscapes, and Civilizations*, pages 75–88. American Geophysical Union (AGU).

Bhushan, R., Sati, S., Rana, N., Shukla, A., Mazumdar, A., and Juyal, N. (2018). High-resolution millennial and centennial scale Holocene monsoon variability in the Higher Central Himalayas. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 489:95 – 104.

Bollasina, M. and Nigam, S. (2011). Modeling of regional hydroclimate change over the Indian subcontinent: Impact of the expanding Thar Desert. *Journal of Climate*, 24(12):3089–3106.

Boschi, R. and Lucarini, V. (2019). Water pathways for the Hindu-Kush-Himalaya and an analysis of three flood events. *Atmosphere*, 10(9):489.

Bosmans, J. H. C., Drijfhout, S. S., Tuenter, E., Lourens, L. J., Hilgen, F. J., and Weber, S. L. (2012). Monsoonal response to mid-Holocene orbital forcing in a high resolution GCM. *Climate of the Past*, 8(2):723–740.

Boucher, O., Denvil, S., Caubel, A., and Foujols, M. A. (2018). IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, 12(7).

Braconnot, P., Zhu, D., Marti, O., and Servonnat, J. (2019). Strengths and challenges for transient Mid- to Late Holocene simulations with dynamical vegetation. *Climate of the Past*, 15(3):997–1024.

Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Prentice-Hall, Englewood Cliffs, New Jersey.

Brierley, C. M., Zhao, A., Harrison, S. P., Braconnot, P., Williams, C. J. R., Thornalley, D. J. R., Shi, X., Peterschmitt, J.-Y., Ohgaito, R., Kaufman, D. S., Kageyama, M., Hargreaves, J. C., Erb, M. P., Emile-Geay, J., D'Agostino, R., Chandan, D., Carré, M., Bartlein, P., Zheng, W., Zhang, Z., Zhang, Q., Yang, H., Volodin, E. M., Tomas, R. A., Routson, C., Peltier, W. R., Otto-Bliesner, B., Morozova, P. A., McKay, N. P., Lohmann, G., Legrande, A. N., Guo, C., Cao, J., Brady, E., Annan, J. D., and Abe-Ouchi, A. (2020). Large-scale

features and evaluation of the PMIP4-CMIP6 *midHolocene* simulations. *Climate of the Past Discussions*, 2020:1–35.

Brown, J. R., Brierley, C. M., An, S.-I., Guarino, M.-V., Stevenson, S., Williams, C. J. R., Zhang, Q., Zhao, A., Braconnot, P., Brady, E. C., Chandan, D., D'Agostino, R., Guo, C., LeGrande, A. N., Lohmann, G., Morozova, P. A., Ohgaito, R., O'ishi, R., Otto-Bliesner, B., Peltier, W. R., Shi, X., Sime, L., Volodin, E. M., Zhang, Z., and Zheng, W. (2020). Comparison of past and future simulations of ENSO in CMIP5/PMIP3 and CMIP6/PMIP4 models. *Climate of the Past Discussions*, 2020:1–44.

Bryson, R. and Swain, A. (1981). Holocene variations of monsoon rainfall in Rajasthan. *Quaternary Research*, 16(2):135 – 145.

Burdanowitz, N., Gaye, B., Hilbig, L., Lahajnar, N., Lückge, A., Rixen, T., and Emeis, K.-C. (2019). Holocene monsoon and sea level-related changes of sedimentation in the northeastern Arabian Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, 166:6 – 18. The 2nd International Indian Ocean Expedition (IIOE-2): Motivating New Exploration in a Poorly Understood Ocean Basin (Volume 2).

C, R. (1996). Languages families and the spread of farming. In Harris, D. R. and Wood, B., editors, *The origins and spread of agriculture and pastoralism in Eurasia*, page 70–92. Smithsonian Institution Press.

Cannon, F., Carvalho, L. M., Jones, C., and Bookhagen, B. (2015). Multi-annual variations in winter westerly disturbance activity affecting the Himalaya. *Climate dynamics*, 44(1-2):441–455.

Cannon, F., Carvalho, L. M., Jones, C., and Norris, J. (2016a). Winter westerly disturbance dynamics and precipitation in the western Himalaya and Karakoram: a wave-tracking approach. *Theoretical and Applied Climatology*, 125(1-2):27–44.

Cannon, F., Carvalho, L. M. V., Jones, C., Hoell, A., Norris, J., Kiladis, G. N., and Tahir, A. A. (2016b). The influence of tropical forcing on extreme winter precipitation in the western Himalaya. *Climate Dynamics*, 48(3-4):1213–1232.

Carlson, A. E. (2010). What caused the Younger Dryas cold event? *Geology*, 38(4):383–384.

Carolin, S. A., Walker, R. T., Day, C. C., Ersek, V., Sloan, R. A., Dee, M. W., Talebian, M., and Henderson, G. M. (2019). Precise timing of abrupt increase in dust activity in the Middle East coincident with 4.2 ka social change. *Proceedings of the National Academy of Sciences*, 116(1):67–72.

Casale, F., Bombelli, G. M., Monti, R., and Bocchiola, D. (2019). Hydropower potential in the Kabul River under climate change scenarios in the XXI century. *Theoretical and Applied Climatology*, 139(3-4):1415–1434.

Cassou, C., Terray, L., Hurrell, J. W., and Deser, C. (2004). North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing. *Journal of Climate*, 17(5):1055–1068.

Castellano, E., Becagli, S., Hansson, M., Hutterli, M., Petit, J. R., Rampino, M. R., Severi, M., Steffensen, J. P., Traversi, R., and Udisti, R. (2005). Holocene volcanic history as recorded in the sulfate stratigraphy of the European Project for Ice Coring in Antarctica Dome C (EDC96) ice core. *Journal of Geophysical Research: Atmospheres*, 110(D6).

Chaboureau, J.-P., Chédin, A., and Scott, N. A. (1998). Relationship between sea surface temperature, vertical dynamics, and the vertical distribution of atmospheric water vapor inferred from TOVS observations. *Journal of Geophysical Research: Atmospheres*, 103(D18):23173–23180.

Chakraborty, A., Behera, S. K., Mujumdar, M., Ohba, R., and Yamagata, T. (2006a). Diagnosis of tropospheric moisture over Saudi Arabia and influences of IOD and ENSO. *Monthly Weather Review*, 134(2):598–617.

Chakraborty, S., Bhattacharya, S., Ranhotra, P., Bhattacharyya, A., and Bhushan, R. (2006b). Palaeoclimatic scenario during Holocene around Sangla valley, Kinnaur northwest Himalaya based on multi proxy records. *Current Science*, 91(6).

Chen, F., Yu, Z., Yang, M., Ito, E., Wang, S., Madsen, D. B., Huang, X., Zhao, Y., Sato, T., John B. Birks, H., Boomer, I., Chen, J., An, C., and Wünnemann, B. (2008). Holocene moisture evolution in arid central Asia and its out-of-phase relationship with Asian monsoon history. *Quaternary Science Reviews*, 27(3):351 – 364.

Chen, L., Zheng, W., and Braconnot, P. (2019). Towards understanding the suppressed ENSO activity during mid-Holocene in PMIP2 and PMIP3 simulations. *Climate Dynamics*, 53(1-2):1095–1110.

Chen, S., Lin, B.-Z., Baig, M., Mitra, B., Lopes, R. J., Santos, A. M., Magee, D. A., Azevedo, M., Tarroso, P., Sasazaki, S., Ostrowski, S., Mahgoub, O., Chaudhuri, T. K., Zhang, Y.-p., Costa, V., Royo, L. J., Goyache, F., Luikart, G., Boivin, N., Fuller, D. Q., Mannen, H., Bradley, D. G., and Beja-Pereira, A. (2009). Zebu cattle are an exclusive legacy of the South Asia Neolithic. *Molecular Biology and Evolution*, 27(1):1–6.

Cheng, H., Sinha, A., Verheyden, S., Nader, F. H., Li, X. L., Zhang, P. Z., Yin, J. J., Yi, L., Peng, Y. B., Rao, Z. G., Ning, Y. F., and Edwards, R. L. (2015). The climate variability in northern Levant over the past 20,000 years. *Geophysical Research Letters*, 42(20):8641–8650.

Clift, P. D. and Plumb, R. A. (2008). *The Asian monsoon: causes, history and effects*, volume 288. Cambridge University Press Cambridge.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, , Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J. (2011). The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137(654):1–28.

Cullen, H. M., deMenocal, P. B., Hemming, S., Hemming, G., Brown, F. H., Guilderson, T., and Sirocko, F. (2000). Climate change and the collapse of the Akkadian empire: Evidence from the deep sea. *Geology*, 28(4):379–382.

Czerniawska, J. and Chlachula, J. (2017). Field trip in the Thar Desert: report. *Landform Analysis*, 35.

Dacre, H. F., Hawcroft, M. K., Stringer, M. A., and Hodges, K. I. (2012). An extratropical cyclone atlas: A tool for illustrating cyclone structure and evolution characteristics. *Bulletin of the American Meteorological Society*, 93(10):1497–1502.

Dahri, Z. H., Moors, E., Ludwig, F., Ahmad, S., Khan, A., Ali, I., and Kabat, P. (2018). Adjustment of measurement errors to reconcile precipitation distribution in the high-altitude Indus basin. *International Journal of Climatology*, 38(10):3842–3860.

De, U., Dube, R. K., and Rao, G. P. (2005). Extreme weather events over India in the last 100 years. *J. Ind. Geophys. Union*, 9(3):173–187.

de Vries, A. J., Feldstein, S. B., Riemer, M., Tyrlis, E., Sprenger, M., Baumgart, M., Fnais, M., and Lelieveld, J. (2016). Dynamics of tropical–extratropical interactions and extreme precipitation events in Saudi Arabia in autumn, winter and spring. *Quarterly Journal of the Royal Meteorological Society*, 142(697):1862–1880.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656):553–597.

Demske, D., Tarasov, P. E., Leipe, C., Kotlia, B. S., Joshi, L. M., and Long, T. (2016). Record of vegetation, climate change, human impact and retting of hemp in Garhwal Himalaya (India) during the past 4600 years. *The Holocene*, 26(10):1661–1675.

Demske, D., Tarasov, P. E., Wünnemann, B., and Riedel, F. (2009). Late glacial and Holocene vegetation, Indian monsoon and westerly circulation in the Trans-Himalaya recorded in the lacustrine pollen sequence from Tso Kar, Ladakh, NW India. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 279(3):172 – 185.

Denniston, R. F., González, L. A., Asmerom, Y., Sharma, R. H., and Reagan, M. K. (2000). Speleothem evidence for changes in Indian summer monsoon precipitation over the last 2300 years. *Quaternary Research*, 53(2):196 – 202.

Deotare, B. C., Kajale, M. D., Rajaguru, S. N., Kusumgar, S., Jull, A. J. T., and Donahue, J. D. (2004). Palaeoenvironmental history of Bap-Malar and Kanod playas of western Rajasthan, Thar desert. *Journal of Earth System Science*, 113(3):403–425.

Dimri, A. and Chevuturi, A. (2014). Model sensitivity analysis study for western disturbances over the Himalayas. *Meteorology and Atmospheric Physics*, 123(3-4):155–180.

Dimri, A. and Chevuturi, A. (2016). *Western disturbances-an Indian meteorological perspective*. Springer.

Dimri, A. P. (2004). Impact of horizontal model resolution and orography on the simulation of a western disturbance and its associated precipitation. *Meteorological Applications*, 11(2):115–127.

Dimri, A. P. (2007). The transport of momentum, sensible heat, potential energy and moisture over the western Himalayas during the winter season. *Theoretical and Applied Climatology*, 90(1-2):49–63.

Dimri, A. P. (2013a). Intraseasonal oscillation associated with the Indian winter monsoon. *Journal of Geophysical Research: Atmospheres*, 118(3):1189–1198.

Dimri, A. P. (2013b). Relationship between ENSO phases with Northwest India winter precipitation. *International Journal of Climatology*, 33(8):1917–1923.

Dimri, A. P. and Niyogi, D. (2013). Regional climate model application at subgrid scale on Indian winter monsoon over the western Himalayas. *International Journal of Climatology*, 33(9):2185–2205.

Dimri, A. P., Niyogi, D., Barros, A. P., Ridley, J., Mohanty, U. C., Yasunari, T., and Sikka, D. R. (2015). Western disturbances: A review. *Reviews of Geophysics*, 53(2):225–246.

Dixit, Y., Hodell, D. A., Giesche, A., Tandon, S. K., Gázquez, F., Saini, H. S., Skinner, L. C., Mujtaba, S. A. I., Pawar, V., Singh, R. N., and Petrie, C. A. (2018). Intensified summer monsoon and the urbanization of Indus Civilization in northwest India. *Scientific Reports*, 8(1).

Dixit, Y., Hodell, D. A., and Petrie, C. A. (2014a). Abrupt weakening of the summer monsoon in northwest India 4100 yr ago. *Geology*, 42(4):339–342.

Dixit, Y., Hodell, D. A., Sinha, R., and Petrie, C. A. (2014b). Abrupt weakening of the Indian summer monsoon at 8.2 kyr B.P. *Earth and Planetary Science Letters*, 391:16 – 23.

Doose-Rolinski, H., Rogalla, U., Scheeder, G., Lückge, A., and von Rad, U. (2001). High-resolution temperature and evaporation changes during the Late Holocene in the northeastern Arabian Sea. *Paleoceanography*, 16(4):358–367.

Douville, H., Chauvin, F., and Broqua, H. (2001). Influence of soil moisture on the Asian and African monsoons. Part I: mean monsoon and daily precipitation. *Journal of Climate*, 14(11):2381–2403.

Dugmore, A. J., Cook, G. T., Shore, J. S., Newton, A. J., Edwards, K. J., and Larsen, G. (1995). Radiocarbon dating tephra layers in Britain and Iceland. *Radiocarbon*, 37(2):379–388.

Dutt, S., Gupta, A. K., Wünnemann, B., and Yan, D. (2018). A long arid interlude in the Indian summer monsoon during 4,350 to 3,450 cal. yr BP contemporaneous to the displacement of the Indus valley civilization. *Quaternary International*, 482:83 – 92.

Dyke, A. S., Moore, A., and Robertson, L. (2003). Deglaciation of North America. Geological Survey of Canada, Open File 1574. Thirty-two digital maps at 1:7 000 000 scale with accompanying digital chronological database and one poster (two sheets) with full map series.

Ebert, E. E., Janowiak, J. E., and Kidd, C. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, 88(1):47–64.

Emile-Geay, J. and Tingley, M. (2016). Inferring climate variability from nonlinear proxies: application to palaeo-ENSO studies. *Climate of the Past*, 12(1):31–50.

Ensor, L. A. and Robeson, S. M. (2008). Statistical characteristics of daily precipitation: comparisons of gridded and point datasets. *Journal of Applied Meteorology and Climatology*, 47(9):2468–2476.

Enzel, Y., Ely, L. L., Mishra, S., Ramesh, R., Amit, R., Lazar, B., Rajaguru, S. N., Baker, V. R., and Sandler, A. (1999). High-resolution holocene environmental changes in the Thar Desert, Northwestern India. *Science*, 284(5411):125–128.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development (Online)*, 9(LLNL-JRNL-736881).

Faisal, N. and Gaffar, A. (2012). Development of Pakistan's new area weighted rainfall using Thiessen polygon method. *Pakistan Journal of Meteorology*, 9(17).

Fernandez-Turiel, J. L., Perez–Torrado, F. J., Rodriguez-Gonzalez, A., Saavedra, J., Carracedo, J. C., Rejas, M., Lobo, A., Osterrieth, M., Carrizo, J. I., Esteban, G., Gallardo, J., and Ratto, N. (2019). The large eruption 4.2 ka cal BP in Cerro Blanco, Central Volcanic Zone, Andes: Insights to the Holocene eruptive deposits in the southern Puna and adjacent regions. *Estudios Geológicos*, 75(1):088.

Filippi, L., Palazzi, E., von Hardenberg, J., and Provenzale, A. (2014). Multidecadal variations in the relationship between the NAO and winter precipitation in the Hindu Kush–Karakoram. *Journal of Climate*, 27(20):7890–7902.

Fleitmann, D., Burns, S. J., Mudelsee, M., Neff, U., Kramers, J., Mangini, A., and Matter, A. (2003). Holocene forcing of the Indian monsoon recorded in a stalagmite from southern Oman. *Science*, 300(5626):1737–1739.

Flückiger, J., Monnin, E., Stauffer, B., Schwander, J., Stocker, T. F., Chappellaz, J., Raynaud, D., and Barnola, J.-M. (2002). High-resolution Holocene N2O ice core record and its relationship with CH4 and CO2. *Global Biogeochemical Cycles*, 16(1):10–1–10–8.

Galarneau, Thomas J., J., Hamill, T. M., Dole, R. M., and Perlwitz, J. (2012). A multiscale analysis of the extreme weather events over Western Russia and Northern Pakistan during July 2010. *Monthly Weather Review*, 140(5):1639–1664.

Gardelle, J., Berthier, E., and Arnaud, Y. (2012). Slight mass gain of karakoram glaciers in the early twenty-first century. *Nature geoscience*, 5(5):322.

Geirsdóttir, A., Miller, G. H., Andrews, J. T., Harning, D. J., Anderson, L. S., Florian, C., Larsen, D. J., and Thordarson, T. (2019). The onset of neoglaciation in Iceland and the 4.2 ka event. *Climate of the Past*, 15(1):25–40.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14):5419–5454.

Ghulami, M., Babel, M. S., and Singh Shrestha, M. (2017). Evaluation of gridded precipitation datasets for the Kabul Basin, Afghanistan. *International journal of remote sensing*, 38(11):3317–3332.

Giesche, A. (2020). *A Multi-Archive Reconstruction of Holocene Summer and Winter Monsoon Variability in NW South Asia*. PhD thesis, University of Cambridge.

Giesche, A., Staubwasser, M., Petrie, C. A., and Hodell, D. A. (2019). Indian winter and summer monsoon strength over the 4.2 ka bp event in foraminifer isotope records from the Indus River delta in the Arabian Sea. *Climate of the Past*, 15(1):73–90.

Giosan, L., Clift, P. D., Macklin, M. G., Fuller, D. Q., Constantinescu, S., Durcan, J. A., Stevens, T., Duller, G. A. T., Tabrez, A. R., Gangal, K., Adhikari, R., Alizai, A., Filip, F., VanLaningham, S., and Syvitski, J. P. M. (2012). Fluvial landscapes of the Harappan civilization. *Proceedings of the National Academy of Sciences*, 109(26):E1688–E1694.

Giosan, L., Orsi, W. D., Coolen, M., Wuchter, C., Dunlea, A. G., Thirumalai, K., Munoz, S. E., Clift, P. D., Donnelly, J. P., Galy, V., and Fuller, D. Q. (2018). Neoglacial climate anomalies and the Harappan metamorphosis. *Climate of the Past*, 14(11):1669–1686.

Glasbey, C. and Nevison, I. (1997). Rainfall modelling using a latent gaussian variable. In *Modelling longitudinal and spatially correlated data*, pages 233–242. Springer.

Goodison, B., Sevruk, B., and Klemm, S. (1989). WMO solid precipitation measurement intercomparison: Objectives, methodology, analysis. *Atmos. Depos*, 179:57–64.

Grams, C. M., Wernli, H., Böttcher, M., Čampa, J., Corsmeier, U., Jones, S. C., Keller, J. H., Lenz, C.-J., and Wiegand, L. (2011). The key role of diabatic processes in modifying the upper-tropospheric wave guide: a North Atlantic case-study. *Quarterly Journal of the Royal Meteorological Society*, 137(661):2174–2193.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

Grömping, U. et al. (2006). Relative importance for linear regression in R: the package relaimpo. *Journal of statistical software*, 17(1):1–27.

Gupta, A. K., Anderson, D. M., and Overpeck, J. T. (2003). Abrupt changes in the Asian southwest monsoon during the Holocene and their links to the North Atlantic Ocean. *Nature*, 421(6921):354–357.

Gurung, D. R., Maharjan, S. B., Shrestha, A. B., Shrestha, M. S., Bajracharya, S. R., and Murthy, M. S. R. (2017). Climate and topographic controls on snow cover dynamics in the Hindu Kush Himalaya. *International Journal of Climatology*, 37(10):3873–3882.

Habib, E., Ciach, G. J., and Krajewski, W. F. (2004). A method for filtering out raingauge representativeness errors from the verification distributions of radar and raingauge rainfall. *Advances in Water Resources*, 27(10):967–980.

Harris, I. and Jones, P. (2017). CRU TS4. 00: Climatic Research Unit (CRU) Time-Series (TS) version 4.00 of high resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2015). *Centre for Environmental Data Analysis*, 25.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. *International journal of climatology*, 34(3):623–642.

Hashmi, M. Z. u. R., Masood, A., Mushtaq, H., Bukhari, S. A. A., Ahmad, B., and Tahir, A. A. (2019). Exploring climate change impacts during the first half of the 21st century on flow regime of the transboundary Kabul River in the Hindukush region. *Journal of Water and Climate Change*. jwc2019094.

Hernández, A., Martin-Puertas, C., Moffa-Sánchez, P., Moreno-Chamarro, E., Ortega, P., Blockley, S., Cobb, K. M., Comas-Bru, L., Giralt, S., Goosse, H., Luterbacher, J., Martrat, B., Muscheler, R., Parnell, A., Pla-Rabes, S., Sjolte, J., Scaife, A. A., Swingedouw, D., Wise, E., and Xu, G. (2020). Modes of climate variability: Synthesis and review of proxy-based reconstructions through the Holocene. *Earth-Science Reviews*, 209:103286.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Munoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J.-N., Vitart, F., Yang, X., Zsótér, E., and Zuo, H. (2018). Operational global reanalysis: progress, future directions and synergies with WP. *ERA-report*, Serie 27.

Hewitt, K. (2011). Glacier Change, Concentration, and Elevation Effects in the Karakoram Himalaya, Upper Indus Basin. *Mountain Research and Development*, 31(3):188 – 200.

Hoffman, J. S., Carlson, A. E., Winsor, K., Klinkhammer, G. P., LeGrande, A. N., Andrews, J. T., and Strasser, J. C. (2012). Linking the 8.2 ka event and its freshwater forcing in the Labrador Sea. *Geophysical Research Letters*, 39(18).

Hong, Y., Hong, B., Lin, Q., Shibata, Y., Hirota, M., Zhu, Y., Leng, X., Wang, Y., Wang, H., and Yi, L. (2005). Inverse phase oscillations between the East Asian and Indian Ocean

summer monsoons during the last 12000 years and paleo-El Niño. *Earth and Planetary Science Letters*, 231(3):337 – 346.

Hou, J., D'Andrea, W. J., Wang, M., He, Y., and Liang, J. (2017). Influence of the Indian monsoon and the subtropical jet on climate change on the Tibetan Plateau since the late Pleistocene. *Quaternary Science Reviews*, 163:84 – 94.

Houze, R. A. (2012). Orographic effects on precipitating clouds. *Reviews of Geophysics*, 50(1).

Houze, R. A., Wilton, D. C., and Smull, B. F. (2007). Monsoon convection in the Himalayan region as seen by the TRMM Precipitation Radar. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 133(627):1389–1411.

Huffman, G. J. and Bolvin, D. T. (2013). Version 1.2 GPCP one-degree daily precipitation data set documentation. *NASA, Goddard Space Flight Center, Greenbelt, MD, USA https://rda. ucar. edu/datasets/ds728*, 3.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hydrometeorology*, 8(1):38–55.

Hunt, K. M., Turner, A. G., and Shaffrey, L. C. (2018a). The evolution, seasonality and impacts of western disturbances. *Quarterly Journal of the Royal Meteorological Society*, 144(710):278–290.

Hunt, K. M., Turner, A. G., and Shaffrey, L. C. (2018b). Extreme daily rainfall in Pakistan and north India: scale interactions, mechanisms, and precursors. *Monthly Weather Review*, 146(4):1005–1022.

Hunt, K. M. R. and Turner, A. G. (2019). The role of the subtropical jet in deficient winter precipitation across the Mid-Holocene Indus basin. *Geophysical Research Letters*, 46(10):5452–5459.

Hunt, K. M. R., Turner, A. G., and Shaffrey, L. C. (2019). Representation of western disturbances in CMIP5 models. *Journal of Climate*, 32(7):1997–2011.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, 269(5224):676–679.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M. (2013). An overview of the North Atlantic Oscillation. In *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, pages 1–35. American Geophysical Union (AGU).

Hussain, S., Song, X., Ren, G., Hussain, I., Han, D., and Zaman, M. (2017). Evaluation of gridded precipitation data in the Hindu Kush–Karakoram–Himalaya mountainous area. *Hydrological sciences journal*, 62(14):2393–2405.

Hussain, S., Xianfang, S., Hussain, I., Jianrong, L., Dong Mei, H., Li Hu, Y., and Huang, W. (2015a). Controlling factors of the stable isotope composition in the precipitation of Islamabad, Pakistan. *Advances in Meteorology*, 2015.

Hussain, S., Xianfang, S., Hussain, I., Jianrong, L., Dong Mei, H., Li Hu, Y., and Huang, W. (2015b). Controlling factors of the stable isotope composition in the precipitation of Islamabad, Pakistan. *Advances in Meteorology*, 2015.

Hébert, R., Rehfeld, K., and Laepple, T. (in press). Comparing estimation techniques for timescale-dependent scaling of climate variability in palaeoclimate time series.

Immerzeel, W., Droogers, P., de Jong, S., and Bierkens, M. (2009). Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing. *Remote Sensing of Environment*, 113(1):40 – 49.

Immerzeel, W., Wanders, N., Lutz, A., Shea, J., and Bierkens, M. (2015). Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. *Hydrology and Earth System Sciences*, 19(11):4673–4687.

Iqbal, M. F. and Athar, H. (2018). Validation of satellite based precipitation over diverse topography of Pakistan. *Atmospheric Research*, 201:247–260.

Jeelani, G., Deshpande, R. D., Galkowski, M., and Rozanski, K. (2018). Isotopic composition of daily precipitation along the southern foothills of the Himalayas: impact of marine and continental sources of atmospheric moisture. *Atmospheric Chemistry and Physics*, 18(12):8789–8805.

Jiang, D., Tian, Z., and Lang, X. (2014). Mid-Holocene global monsoon area and precipitation from PMIP simulations. *Climate Dynamics*, 44(9-10):2493–2512.

Johnsen, S. J., Dahl-Jensen, D., Gundestrup, N., Steffensen, J. P., Clausen, H. B., Miller, H., Masson-Delmotte, V., Sveinbjörnsdottir, A. E., and White, J. (2001). Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP. *Journal of Quaternary Science*, 16(4):299–307.

Joshi, L. M., Kotlia, B. S., Ahmad, S. M., Wu, C.-C., Sanwal, J., Raza, W., Singh, A. K., Shen, C.-C., Long, T., and Sharma, A. K. (2017). Reconstruction of Indian monsoon precipitation variability between 4.0 and 1.6 ka BP using speleothem δ 18O records from the Central Lesser Himalaya, India. *Arabian Journal of Geosciences*, 10(16).

Joussaume, S. and Braconnot, P. (1997). Sensitivity of paleoclimate simulation results to season definitions. *Journal of Geophysical Research: Atmospheres*, 102(D2):1943–1956.

Kahkashan, S., Chen, J., Wang, X., Clift, P. D., Jalali, B., Inam, A., Saleem, M., Aftab, J., Ji, Z., Bai, Y., Li, Z., Haiyan, J., Li, H., and Ran, L. (2020). Holocene organic geochemical record from the Western Indus continental shelf (northern Arabian Sea). *The Holocene*, page 0959683620902213.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne,

R., and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–472.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., and Potter, G. (2002). NCEP–DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*, 83(11):1631–1644.

Kaniewski, D., Marriner, N., Cheddadi, R., Guiot, J., and Van Campo, E. (2018). The 4.2 ka BP event in the Levant. *Climate of the Past*, 14(10):1529–1542.

Kar, S. C. and Rana, S. (2013). Interannual variability of winter precipitation over northwest India and adjoining region: impact of global forcings. *Theoretical and Applied Climatology*, 116(3-4):609–623.

Karim, A. and Veizer, J. (2002). Water balance of the Indus River Basin and moisture source in the Karakoram and western Himalayas: Implications from hydrogen and oxygen isotopes in river water. *Journal of Geophysical Research: Atmospheres*, 107(D18):ACH 9–1–ACH 9–12.

Kathayat, G., Cheng, H., Sinha, A., Berkelhammer, M., Zhang, H., Duan, P., Li, H., Li, X., Ning, Y., and Edwards, R. L. (2018). Evaluating the timing and structure of the 4.2 ka event in the Indian summer monsoon domain from an annually resolved speleothem record from Northeast India. *Climate of the Past*, 14(12):1869–1879.

Kathayat, G., Cheng, H., Sinha, A., Spötl, C., Edwards, R. L., Zhang, H., Li, X., Yi, L., Ning, Y., Cai, Y., Lui, L. W., and Breitenbach, S. F. M. (2016). Indian monsoon variability on millennial-orbital timescales. *Scientific Reports*, 6(1).

Kathayat, G., Cheng, H., Sinha, A., Yi, L., Li, X., Zhang, H., Li, H., Ning, Y., and Edwards, R. L. (2017). The Indian monsoon variability and civilization changes in the Indian subcontinent. *Science Advances*, 3(12).

Kaushal, N., Breitenbach, S. F. M., Lechleitner, F. A., Sinha, A., Tewari, V. C., Ahmad, S. M., Berkelhammer, M., Band, S., Yadava, M., Ramesh, R., and et al. (2018). The Indian summer monsoon from a speleothem δ 18o perspective—a review. *Quaternary*, 1(3):29.

Kenoyer, J. M. (1997). Trade and technology of the Indus Valley: New insights from Harappa, Pakistan. *World Archaeology*, 29(2):262–280.

Kenoyer, J. M. (2008). Indus urbanism: New perspectives on its origin and character. In Marcus, J. and Sabloff, J. A., editors, *The ancient city: New perspectives on urbanism in the old and new world*, pages 183–208. School for Advanced Research Press Santa Fe, NM.

Khan, A., Hameed, A., Kumar, A., Ibrar, N., Iqbal, Zafar nd Iqbal, M. M., and Amin, M. (2011). Annual flood report 2010. Technical report, Federal Flood Commission.

Khan, S. I., Hong, Y., Gourley, J. J., Khattak, M. U. K., Yong, B., and Vergara, H. J. (2014). Evaluation of three high-resolution satellite precipitation estimates: Potential for monsoon monitoring over Pakistan. *Advances in Space Research*, 54(4):670–684.

Kishore, P., Jyothi, S., Basha, G., Rao, S., Rajeevan, M., Velicogna, I., and Sutterley, T. C. (2016). Precipitation climatology over India: validation with observations and reanalysis datasets and spatial trends. *Climate dynamics*, 46(1-2):541–556.

Klein, S. A., Soden, B. J., and Lau, N.-C. (1999). Remote sea surface temperature variations during ENSO: Evidence for a tropical atmospheric bridge. *Journal of Climate*, 12(4):917–932.

Kobashi, T., Menviel, L., Jeltsch-Thömmes, A., Vinther, B. M., Box, J. E., Muscheler, R., Nakaegawa, T., Pfister, P. L., Döring, M., Leuenberger, M., Wanner, H., and Ohmura, A. (2017). Volcanic influence on centennial to millennial Holocene Greenland temperature change. *Scientific Reports*, 7(1).

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1):5–48.

Kothawale, D., Munot, A., and Kumar, K. K. (2010). Surface air temperature variability over India during 1901–2007, and its association with ENSO. *Climate Research*, 42(1):89–104.

Kotlia, B. S., Singh, A. K., Joshi, L. M., and Bisht, K. (2018). Precipitation variability over Northwest Himalaya from 4.0 to 1.9 ka BP with likely impact on civilization in the foreland areas. *Journal of Asian Earth Sciences*, 162:148 – 159. Emergence and Evolution of Himalayan Foreland Basin.

Kotlia, B. S., Singh, A. K., Joshi, L. M., and Dhaila, B. S. (2015). recipitation variability in the Indian Central Himalaya during last ca. 4,000 years inferred from a speleothem record: Impact of Indian Summer Monsoon (ISM) and Westerlies. *Quaternary International*, 371:244 – 253. Updated Quaternary Climatic Research in parts of the Third Pole Selected papers from the HOPE-2013 conference, Nainital, India.

Krishnamurti, T. N. (1961). The subtropical jet stream of winter. *Journal of Meteorology*, 18(2):172–191.

Krishnan, R., Sabin, T. P., Madhura, R. K., Vellore, R. K., Mujumdar, M., Sanjay, J., Nayak, S., and Rajeevan, M. (2018). Non-monsoonal precipitation response over the Western Himalayas to climate change. *Climate Dynamics*, 52(7-8):4091–4109.

Kug, J.-S., Li, T., An, S.-I., Kang, I.-S., Luo, J.-J., Masson, S., and Yamagata, T. (2006). Role of the ENSO–Indian Ocean coupling on ENSO variability in a coupled GCM. *Geophysical Research Letters*, 33(9).

Kutzbach, J. E., Liu, X., Liu, Z., and Chen, G. (2007). Simulation of the evolutionary response of global summer monsoons to orbital forcing over the past 280,000 years. *Climate Dynamics*, 30(6):567–579.

Lachniet, M. S. (2009). Climatic and environmental controls on speleothem oxygen-isotope values. *Quaternary Science Reviews*, 28(5):412 – 432.

Lang, T. J. and Barros, A. P. (2004). Winter storms in the central Himalayas. *Journal of the Meteorological Society of Japan. Ser. II*, 82(3):829–844.

Lau, W. K. M. and Kim, K.-M. (2012). The 2010 Pakistan flood and Russian heat wave: Teleconnection of hydrometeorological extremes. *Journal of Hydrometeorology*, 13(1):392–403.

Law, R. (2006). Moving mountains: The trade and transport of rocks and minerals with in the greater Indus Valley region. In EC Robertson, RD Seibert, D. F. and Zender, M., editors, *Space and Spatial Analysis in Archaeology*. University of Calgary Press, Alberta, Canada.

Law, R. W. (2011). *Inter-regional interaction and urbanism in the ancient Indus Valley: A geologic provenience study of Harappa's rock and mineral assemblage*. Indus Project, Research Institute for Humanity and Nature.

Legates, D. R. and Willmott, C. J. (1990). Mean seasonal and spatial variability in gauge-corrected, global precipitation. *International Journal of Climatology*, 10(2):111–127.

LeGrande, A. N. and Schmidt, G. A. (2006). Global gridded data set of the oxygen isotopic composition in seawater. *Geophysical Research Letters*, 33(12).

Leipe, C., Demske, D., and Tarasov, P. E. (2014). A Holocene pollen record from the north-western Himalayan lake Tso Moriri: Implications for palaeoclimatic and archaeological research. *Quaternary International*, 348:93 – 112. The Bridging Eurasia Research Initiative: Modes of mobility and sustainability in the palaeoenvironmental and archaeological archives from Eurasia.

Liu, Z., Harrison, S. P., Kutzbach, J., and Otto-Bliesner, B. (2004). Global monsoons in the mid-Holocene and oceanic feedback. *Climate Dynamics*, 22(2-3):157–182.

Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S., Lohmann, G., Zheng, W., and Elison Timm, O. (2014). The Holocene temperature conundrum. *Proceedings of the National Academy of Sciences*, 111(34):E3501–E3505.

Lovejoy, S. and Schertzer, D. (2012). Haar wavelets, fluctuations and structure functions: convenient choices for geophysics. *Nonlinear Processes in Geophysics*, 19(5):513–527.

Lückge, A., Doose-Rolinski, H., Khan, A., Schulz, H., and von Rad, U. (2001). Monsoonal variability in the northeastern Arabian Sea during the past 5000years: geochemical evidence from laminated sediments. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 167(3):273 – 286.

Madhura, R. K., Krishnan, R., Revadekar, J. V., Mujumdar, M., and Goswami, B. N. (2015). Changes in western disturbances over the Western Himalayas in a warming environment. *Climate Dynamics*, 44(3-4):1157–1168.

Mahanta, R., Sarma, D., and Choudhury, A. (2013). Heavy rainfall occurrences in northeast India. *International Journal of Climatology*, 33(6):1456–1469.

Malardel, S. (2005). *Fondamentaux de météorologie: à l'école du temps*. Cépaduès.

Malurkar, S. L. (1947). Abnormally dry and wet western disturbances over North India. *Current Science*, 16(5):139–141.

Mann, M. E., Steinman, B. A., and Miller, S. K. (2020). Absence of internal multidecadal and interdecadal oscillations in climate model simulations. *Nature Communications*, 11(1).

Mannan, M. A., Chowdhury, M. A. M., Karmakar, S., Ahmed, S., and Rahman, A. (2017). Analysis and prediction of rainfall associated with western disturbances during winter months in Bangladesh. *Den-Drop (A Scientific Journal of Meteorology and Geo-physics), Bangladesh Meteorological Department, Dhaka, Bangladesh*, 4:12–24.

Mantsis, D. F., Clement, A. C., Broccoli, A. J., and Erb, M. P. (2011). Climate Feedbacks in Response to Changes in Obliquity. *Journal of Climate*, 24(11):2830–2845.

Mariotti, A. (2007). How ENSO impacts precipitation in southwest central Asia. *Geophysical Research Letters*, 34(16).

Martínez-Alvarado, O., Joos, H., Chagnon, J., Boettcher, M., Gray, S. L., Plant, R. S., Methven, J., and Wernli, H. (2014). The dichotomous structure of the warm conveyor belt. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1809–1824.

Masato, G., Hoskins, B. J., and Woollings, T. J. (2012). Wave-breaking characteristics of mid-latitude blocking. *Quarterly Journal of the Royal Meteorological Society*, 138(666):1285–1296.

Matero, I., Gregoire, L., Ivanovic, R., Tindall, J., and Haywood, A. (2017). The 8.2 ka cooling event caused by Laurentide ice saddle collapse. *Earth and Planetary Science Letters*, 473:205 – 214.

Medina, S., Houze, R. A., Kumar, A., and Niyogi, D. (2010). Summer monsoon convection in the Himalayan region: Terrain and land cover effects. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 136(648):593–616.

Meher, J. K., Das, L., Akhter, J., Benestad, R. E., and Mezghani, A. (2017). Performance of CMIP3 and CMIP5 GCMs to simulate observed rainfall characteristics over the Western Himalayan region. *Journal of Climate*, 30(19):7777–7799.

Ménégoz, M., Gallée, H., and Jacobi, H. (2013). Precipitation and snow cover in the Himalaya: from reanalysis to regional climate simulations. *Hydrology and Earth System Sciences*, 17(10):3921–3936.

Michel, C., Rivière, G., Terray, L., and Joly, B. (2012). The dynamical link between surface cyclones, upper-tropospheric Rossby wave breaking and the life cycle of the Scandinavian blocking. *Geophysical Research Letters*, 39(10).

Midhuna, T. M., Kumar, P., and Dimri, A. P. (2020). A new western disturbance index for the Indian winter monsoon. *Journal of Earth System Science*, 129(1).

Mishra, P. K., Anoop, A., Schettler, G., Prasad, S., Jehangir, A., Menzel, P., Naumann, R., Yousuf, A., Basavaiah, N., Deenadayalan, K., Wiesner, M., and Gaye, B. (2015). Reconstructed late Quaternary hydrological changes from Lake Tso Moriri, NW Himalaya. *Quaternary International*, 371:76 – 86. Updated Quaternary Climatic Research in parts of the Third Pole Selected papers from the HOPE-2013 conference, Nainital, India.

Misra, P., Tandon, S., and Sinha, R. (2019). Holocene climate records from lake sediments in India: Assessment of coherence across climate zones. *Earth-Science Reviews*, 190:370 – 397.

Muhammad Tahir, K., Yin, Y., Wang, Y., Babar, Z. A., and Yan, D. (2015). Impact assessment of orography on the extreme precipitation event of July 2010 over Pakistan: a numerical study. *Advances in Meteorology*, 2015.

Mujumdar, M. (2006). *Diagnostic Analysis of Wintertime Rainfall Events Over the Arabian Region*. Contribution from IITM. Indian Institute of Tropical Meteorology.

Mull, S. and Desai, B. (1947). The origin and structure of the winter depressions of Northwest India. Technical Note 25:18.

Murakami, M. (1995). *Managing water for peace in the Middle East: alternative strategies*. United Nations University Press.

Nakamura, A., Yokoyama, Y., Maemoku, H., Yagi, H., Okamura, M., Matsuoka, H., Miyake, N., Osada, T., Adhikari, D. P., Dangol, V., Ikehara, M., Miyairi, Y., and Matsuzaki, H. (2016). Weak monsoon event at 4.2 ka recorded in sediment from Lake Rara, Himalayas. *Quaternary International*, 397:349 – 359. Japanese Quaternary Studies.

NASA/GISS (2018). NASA-GISS GISS-E2.1G model output prepared for CMIP6 CMIP historical.

Nathans, L., Oswald, F., and Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research and Evaluation*, 17(9):1–19.

Newlands, N. K., Davidson, A., Howard, A., and Hill, H. (2011). Validation and inter-comparison of three methodologies for interpolating daily precipitation and temperature across Canada. *Environmetrics*, 22(2):205–223.

Oman, L., Robock, A., Stenchikov, G., Schmidt, G. A., and Ruedy, R. (2005). Climatic response to high-latitude volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, 110(D13).

Orengo, H. and Petrie, C. (2017). Large-scale, multi-temporal remote sensing of palaeo-river networks: A case study from northwest India and its implications for the Indus Civilisation. *Remote Sensing*, 9(7):735.

Orengo, H. A. and Petrie, C. A. (2018). Multi-scale relief model (MSRM): a new algorithm for the visualization of subtle topographic change of variable size in digital elevation models. *Earth Surface Processes and Landforms*, 43(6):1361–1369.

Otto-Bliesner, B. L., Braconnot, P., Harrison, S. P., Lunt, D. J., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Capron, E., Carlson, A. E., Dutton, A., Fischer, H., Goelzer, H., Govin, A., Haywood, A., Joos, F., LeGrande, A. N., Lipscomb, W. H., Lohmann, G., Mahowald, N., Nehrbass-Ahles, C., Pausata, F. S. R., Peterschmitt, J.-Y., Phipps, S. J., Renssen, H., and Zhang, Q. (2017). The PMIP4 contribution to CMIP6 – Part 2: Two interglacials, scientific objective and experimental design for Holocene and Last Interglacial simulations. *Geoscientific Model Development*, 10(11):3979–4003.

Palazzi, E., Hardenberg, J. V., Terzago, S., and Provenzale, A. (2015). Precipitation in the Karakoram-Himalaya: a CMIP5 view. *Climate Dynamics*, 45(1-2):21–45.

Palazzi, E., Von Hardenberg, J., and Provenzale, A. (2013). Precipitation in the Hindu-Kush Karakoram Himalaya: observations and future scenarios. *Journal of Geophysical Research: Atmospheres*, 118(1):85–100.

Paliwal, B. S. (2013). Abrupt Holocene climatic change in Northwestern India: Disappearance of the Sarasvati River and the end of Vedic Civilization. In *Abrupt Climate Change: Mechanisms, Patterns, and Impacts*, pages 185–194. American Geophysical Union (AGU).

Parker, A., Davies, C., and Wilkinson, T. (2006). The early to mid-Holocene moist period in Arabia: some recent evidence from lacustrine sequences in eastern and south-western Arabia. *Proceedings of the Seminar for Arabian Studies*, 36:243–255.

Pathak, A., Ghosh, S., Martinez, J. A., Dominguez, F., and Kumar, P. (01 Mar. 2017). Role of oceanic and land moisture sources and transport in the seasonal and interannual variability of summer monsoon in India. *Journal of Climate*, 30(5):1839 – 1859.

Pausata, F. S., Messori, G., and Zhang, Q. (2016). Impacts of dust reduction on the northward expansion of the African monsoon during the Green Sahara period. *Earth and Planetary Science Letters*, 434:298 – 307.

Pelly, J. L. and Hoskins, B. J. (2003). A new perspective on blocking. *Journal of the Atmospheric Sciences*, 60(5):743–755.

Perșoiu, A., Onac, B. P., Wynn, J. G., Blaauw, M., Ionita, M., and Hansson, M. (2017). Holocene winter climate variability in Central and Eastern Europe. *Scientific Reports*, 7(1).

Petrie, C., Parikh, D., Green, A., and Bates, J. (2018). Looking beneath the veneer: thoughts about environmental and cultural diversity in the Indus civilisation. In Frenez, D., Jamison, G., Law, R., Vidale, M., and Meadow, R., editors, *Walking with the Unicorn, Social Organization and Material Culture in Ancient South Asia*, pages 455–476. Archaeopress, Oxford, UK.

Petrie, C. and Thomas, K. (2012). The topographic and environmental context of the earliest village sites in western South Asia. *Antiquity*, 86(334):1055–1067.

Petrie, C. A. (2013). South Asia. *Oxford Handbooks Online*.

Petrie, C. A. (2015). Mehrgarh, Pakistan. In Barker, G. and Goucher, C., editors, *The Cambridge World History*, volume 2 of *The Cambridge World History*, page 289–309. Cambridge University Press.

Petrie, C. A. (2017). Crisis, what crisis? Adaptation, resilience and transformation in the Indus Civilisation. In Cunningham, T. and Driessen, J., editors, *Crisis to Collapse: The Archaeology of Social Breakdown*. Aegis Publications, UC Louvain.

Petrie, C. A. (2019). Diversity, variability, adaptation and 'fragility'in the indus civilization. In Yoffee, N., editor, *The Evolution of Fragility: Setting the Terms*, chapter 7. McDonald Institute Monographs: Conversations.

Petrie, C. A. and Bates, J. (2017). 'Multi-cropping', Intercropping and Adaptation to Variable Environments in Indus South Asia. *Journal of World Prehistory*, 30(2):81–130.

Petrie, C. A., Singh, R. N., Bates, J., Dixit, Y., French, C. A., Hodell, D. A., Jones, P. J., Lancelotti, C., Lynam, F., Neogi, S., et al. (2017). Adaptation to variable environments, resilience to climate change: Investigating land, water and settlement in Indus Northwest India. *Current Anthropology*, 58(1):000–000.

Petrie, C. A. and Weeks, L. R. (2019). The Iranian Plateau and the Indus River basin. In Chiotis, E., editor, *Climate Changes in the Holocene:Their Impacts and Human Adaptation*, pages 293–326. Taylor & Francis, Boca Raton.

Phadtare, N. R. (2000). Sharp decrease in summer monsoon strength 4000–3500 cal yr b.p. in the central higher Himalaya of India based on pollen evidence from alpine peat. *Quaternary Research*, 53(1):122 – 129.

Pisharoty, P. and Desai, B. (1956). Western disturbances and Indian weather. *Indian J. Meteorol. Geophys*, 7:333–338.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11):4083–4097.

Possehl, G. L. (1997a). Climate and the eclipse of the ancient cities of the Indus. *Third Millennium BC Climate Change and Old World Collapse*, page 193–243.

Possehl, G. L. (1997b). The transformation of the Indus civilization. *Journal of World Prehistory*, 11(4):425–472.

Possehl, G. L. (1998). Sociocultural complexity without the state: the Indus civilization. *Archaic states*, pages 261–291.

Possehl, G. L. (1999). *Indus Age: the beginnings*, volume 1. University of Pennsylvania Press.

Possehl, G. L. (2002). *The Indus civilization: a contemporary perspective*. Rowman Altamira.

Prasad, S., Anoop, A., Riedel, N., Sarkar, S., Menzel, P., Basavaiah, N., Krishnan, R., Fuller, D., Plessen, B., Gaye, B., Röhl, U., Wilkes, H., Sachse, D., Sawant, R., Wiesner, M., and Stebich, M. (2014a). Prolonged monsoon droughts and links to Indo-Pacific warm pool: A Holocene record from Lonar Lake, central India. *Earth and Planetary Science Letters*, 391:171 – 182.

Prasad, S. and Enzel, Y. (2006). Holocene paleoclimates of India. *Quaternary Research*, 66(3):442 – 453. Holocene Climate and Cultural Evolution in Late Prehistoric-Early Historic West Asia.

Prasad, S., Kusumgar, S., and Gupta, S. K. (1997). A mid to late Holocene record of palaeoclimatic changes from Nal Sarovar: a palaeodesert margin lake in western India. *Journal of Quaternary Science*, 12(2):153–159.

Prasad, S., Mishra, P. K., Menzel, P., Gaye, B., Jehangir, A., and Yousuf, A. R. (2016). esting the validity of productivity proxy indicators in high altitude Tso Moriri Lake, NW Himalaya (India). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 449:421 – 430.

Prasad, V., Farooqui, A., Sharma, A., Phartiyal, B., Chakraborty, S., Bhandari, S., Raj, R., and Singh, A. (2014b). Mid–late Holocene monsoonal variations from mainland Gujarat, India: A multi-proxy study for evaluating climate culture relationship. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 397:38 – 51. SI: Quaternary Monsoon History from India and adjoining seas.

Prasad, V., Phartiyal, B., and Sharma, A. (2007). Evidence of enhanced winter precipitation and the prevalence of a cool and dry climate during the mid to late Holocene in mainland Gujarat, India. *The Holocene*, 17(7):889–896.

Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In *Proceedings of the second international Tampere conference in statistics, 1987*, pages 245–260. Department of Mathematical Sciences, University of Tampere.

Rahman, S., Sengupta, D., and Ravichandran, M. (2009). Variability of Indian summer monsoon rainfall in daily data from gauge and satellite. *Journal of Geophysical Research: Atmospheres*, 114(D17).

Railsback, L. B., Liang, F., Brook, G., Voarintsoa, N. R. G., Sletten, H. R., Marais, E., Hardt, B., Cheng, H., and Edwards, R. L. (2018). The timing, two-pulsed nature, and variable climatic expression of the 4.2 ka event: A review and new high-resolution stalagmite data from Namibia. *Quaternary Science Reviews*, 186:78 – 90.

Raj, R., Chamyal, L., Prasad, V., Sharma, A., Tripathi, J. K., and Verma, P. (2015). Holocene climatic fluctuations in the Gujarat Alluvial Plains based on a multiproxy study of the Pariyaj Lake archive, western India. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 421:60 – 74.

Rajbhandari, R., Shrestha, A., Kulkarni, A., Patwardhan, S., and Bajracharya, S. (2015). Projected changes in climate over the Indus river basin using a high resolution regional climate model (PRECIS). *Climate Dynamics*, 44(1-2):339–357.

Rakesh, V., Singh, R., Yuliya, D., Pal, P. K., and Joshi, P. C. (2009). Impact of variational assimilation of MODIS thermodynamic profiles in the simulation of western disturbance. *International Journal of Remote Sensing*, 30(18):4867–4887.

Rana, S., McGregor, J., and Renwick, J. (2015). Precipitation seasonality over the Indian subcontinent: An evaluation of gauge, reanalyses, and satellite retrievals. *Journal of Hydrometeorology*, 16(2):631–651.

Rana, S., McGregor, J., and Renwick, J. (2017). Wintertime precipitation climatology and ENSO sensitivity over central southwest asia. *International Journal of Climatology*, 37(3):1494–1509.

Rawat, S., Gupta, A. K., Sangode, S., Srivastava, P., and Nainwal, H. (2015). Late Pleistocene–Holocene vegetation and Indian summer monsoon record from the Lahaul, Northwest Himalaya, India. *Quaternary Science Reviews*, 114:167 – 181.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108(D14).

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P., Koster, R. D., and De Lannoy, G. J. (2017). Assessment of MERRA-2 land surface hydrology estimates. *Journal of Climate*, 30(8):2937–2960.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743):195–204.

Rex, D. F. (1950). Blocking action in the middle troposphere and its effect upon regional climate. *Tellus*, 2(3):196–211.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of climate*, 24(14):3624–3648.

Robock, A. (2000). Volcanic eruptions and climate. *Reviews of Geophysics*, 38(2):191–219.

Roy, P., Nagar, Y., Juyal, N., Smykatz-Kloss, W., and Singhvi, A. (2009). Geochemical signatures of Late Holocene paleo-hydrological changes from Phulera and Pokharan saline playas near the eastern and western margins of the Thar Desert, India. *Journal of Asian Earth Sciences*, 34(3):275 – 286.

Roy, P. D. and Singhvi, A. K. (2016). Climate variation in the Thar Desert since the Last Glacial Maximum and evaluation of the Indian Monsoon. *TIP*, 19(1):32 – 44.

Rudolf, B. and Rubel, F. (2005). Global precipitation. In *Observed global climate*, chapter 11, pages 1–24. Springer.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1058.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., and Becker, E. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, 27(6):2185–2208.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751):360–363.

Sandeep, K., Shankar, R., Warrier, A. K., Yadava, M., Ramesh, R., Jani, R., Weijian, Z., and Xuefeng, L. (2017). A multi-proxy lake sediment record of Indian summer monsoon variability during the Holocene in southern India. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 476:1 – 14.

Saporta, G. (2006). *Probabilités, analyse des données et statistique*. Editions Technip.

Scherrer, S. C., Croci-Maspoli, M., Schwierz, C., and Appenzeller, C. (2006). Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *International Journal of Climatology*, 26(2):233–249.

Schiemann, R., Lüthi, D., and Schär, C. (2009). Seasonality and interannual variability of the westerly jet in the Tibetan Plateau region. *Journal of Climate*, 22(11):2940–2957.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., and Ziese, M. (2018). GPCC full data monthly product version 2018 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historical data. *Deutscher Wetterdienst: Offenbach am Main, Germany*.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B. (2014). GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, 115(1-2):15–40.

Sevruk, B. (1984). International comparison of national precipitation gauges with a reference pit gauge. *WMO Instrument and Observing Methods Report, No. 17*, 111.

Singh, G., Joshi, R., and Singh, A. (1972). Stratigraphic and radiocarbon evidence for the age and development of three salt lake deposits in Rajasthan, India. *Quaternary Research*, 2(4):496 – 505.

Singh, G., Joshi, R. D., Chopra, S. K., Singh, A. B., and West, R. G. (1974). Late quaternary history of vegetation and climate of the Rajasthan desert, India. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 267(889):467–501.

Singh, G., Wasson, R., and Agrawal, D. (1990). Vegetational and seasonal climatic changes since the last full glacial in the Thar Desert, northwestern India. *Review of Palaeobotany and Palynology*, 64(1):351 – 358. The Proceedings of the 7th International Palynological Congress (Part I).

Singh, P. and Jain, S. K. (2002). Snow and glacier melt in the Satluj River at Bhakra Dam in the western Himalayan region. *Hydrological Sciences Journal*, 47(1):93–106.

Singh, S. P., Bassignana-Khadka, I., Singh Karky, B., and Sharma, E. (2011). Climate change in the Hindu Kush-Himalayas: the state of current knowledge. Technical report, International Centre for Integrated Mountain Development (ICIMOD).

Singhvi, A. K. and Kar, A. (2004). The aeolian sedimentation record of the Thar desert. *Journal of Earth System Science*, 113(3):371–401.

Sinha, A., Kathayat, G., Cheng, H., Breitenbach, S. F. M., Berkelhammer, M., Mudelsee, M., Biswas, J., and Edwards, R. L. (2015). Trends and oscillations in the Indian summer monsoon rainfall over the last two millennia. *Nature Communications*, 6(1).

Sinha, R. (2009). The great avulsion of Kosi on 18 august 2008. *Current Science*, 97(3):429–433.

Sinha, R., Smykatz-Kloss, W., Stüben, D., Harrison, S., Berner, Z., and Kramar, U. (2006). Late Quaternary palaeoclimatic reconstruction from the lacustrine sediments of the Sambhar playa core, Thar Desert margin, India. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 233(3):252 – 270.

Sirocko, F., Sarnthein, M., Erlenkeuser, H., Lange, H., Arnold, M., and Duplessy, J. C. (1993). Century-scale events in monsoonal climate over the past 24,000 years. *Nature*, 364(6435):322–324.

Smith, R. L., Joel W. Ager, J., and Williams, D. L. (1992). Suppressor variables in multiple regression/correlation. *Educational and Psychological Measurement*, 52(1):17–29.

Srivastava, P., Agnihotri, R., Sharma, D., Meena, N., Sundriyal, Y. P., Saxena, A., Bhushan, R., Sawlani, R., Banerji, U. S., Sharma, C., Bisht, P., Rana, N., and Jayangondaperumal, R. (2017). 8000-year monsoonal record from Himalaya revealing reinforcement of tropical and global climate systems since mid-Holocene. *Scientific Reports*, 7(1).

Staubwasser, M., Sirocko, F., Grootes, P. M., and Erlenkeuser, H. (2002). South Asian monsoon climate change and radiocarbon in the Arabian Sea during early and middle Holocene. *Paleoceanography*, 17(4):15–1–15–12.

Staubwasser, M., Sirocko, F., Grootes, P. M., and Segl, M. (2003). Climate change at the 4.2 ka BP termination of the Indus valley civilization and Holocene south Asian monsoon variability. *Geophysical Research Letters*, 30(8).

Staubwasser, M. and Weiss, H. (2006). Holocene climate and cultural evolution in late prehistoric–early historic West Asia. *Quaternary Research*, 66(3):372–387.

Steinhilber, F., Beer, J., and Fröhlich, C. (2009). Total solar irradiance during the Holocene. *Geophysical Research Letters*, 36(19).

Strong, C. and Magnusdottir, G. (01 Sep. 2008). Tropospheric Rossby wave breaking and the NAO/NAM. *Journal of the Atmospheric Sciences*, 65(9):2861 – 2876.

Swain, A., Kutzbach, J., and Hastenrath, S. (1983). Estimates of Holocene precipitation for Rajasthan, India, based on pollen and lake-level data. *Quaternary Research*, 19(1):1 – 17.

Syed, F. S., Giorgi, F., Pal, J. S., and Keay, K. (2010). Regional climate model simulation of winter climate over Central–Southwest Asia, with emphasis on NAO and ENSO effects. *International Journal of Climatology*, 30(2):220–235.

Syed, F. S., Giorgi, F., Pal, J. S., and King, M. P. (2006). Effect of remote forcings on the winter precipitation of central southwest Asia part 1: observations. *Theoretical and Applied Climatology*, 86(1-4):147–160.

Syvitski, J. P., Kettner, A. J., Overeem, I., Giosan, L., Brakenridge, G. R., Hannon, M., and Bilham, R. (2013). Anthropocene metamorphosis of the Indus delta and lower floodplain. *Anthropocene*, 3:24–35.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Thapa, K., Endreny, T. A., and Ferguson, C. R. (2018). Atmospheric rivers carry nonmonsoon extreme precipitation into Nepal. *Journal of Geophysical Research: Atmospheres*, 123(11):5901–5912.

Thomas, L., Dash, S. K., Mohanty, U. C., and Babu, C. A. (2018). Features of western disturbances simulated over north India using different land-use data sets. *Meteorological Applications*, 25(2):246–253.

Thompson, L. G., Mosley-Thompson, E., Davis, M. E., Henderson, K. A., Brecher, H. H., Zagorodnov, V. S., Mashiotta, T. A., Lin, P.-N., Mikhalenko, V. N., Hardy, D. R., and Beer, J. (2002). Kilimanjaro ice core records: Evidence of holocene climate change in tropical africa. *Science*, 298(5593):589–593.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.

Tibaldi, S. and Molteni, F. (1990). On the operational predictability of blocking. *Tellus A*, 42(3):343–365.

Tinmaker, M. and Ali, K. (2012). Space time variation of lightning activity over northeast India. *Meteorologische Zeitschrift*, 21(2):135–143.

Tozer, C., Kiem, A., and Verdon-Kidd, D. (2012). On the uncertainties associated with using gridded rainfall data as a proxy for observed. *Hydrology and Earth System Sciences*, 16(5):1481–1499.

Trenberth, K. E. (1997). The definition of El Niño. *Bulletin of the American Meteorological Society*, 78(12):2771–2778.

Trenberth, K. E. and Stepaniak, D. P. (2001). Indices of El Niño evolution. *Journal of Climate*, 14(8):1697–1701.

Trivedi, A., Chauman, M., and Malik, M. (2013). Holocene vegetation and climate change in Jammu region, based on pollen evidence from the lake deposits. *Man and Environment*, 38(1):74–89.

Turner, T. E., Swindles, G. T., Charman, D. J., Langdon, P. G., Morris, P. J., Booth, R. K., Parry, L. E., and Nichols, J. E. (2016). Solar cycles or random processes? Evaluating solar variability in Holocene climate records. *Scientific Reports*, 6(1).

Tustison, B., Harris, D., and Foufoula-Georgiou, E. (2001). Scale issues in verification of precipitation forecasts. *Journal of Geophysical Research: Atmospheres*, 106(D11):11775–11784.

ul Islam, M. Z. and Rahmani, A. R. (2011). Thar Desert, Rajasthan, India: anthropogenic influence on biodiversity and grasslands. *Biodiversity*, 12(2):75–89.

Varotsos, C. A., Franzke, C. L. E., Efstathiou, M. N., and Degermendzhi, A. G. (2013). Evidence for two abrupt warming events of SST in the last century. *Theoretical and Applied Climatology*, 116(1-2):51–60.

Vellore, R. K., Kaplan, M. L., Krishnan, R., Lewis, J. M., Sabade, S., Deshpande, N., Singh, B. B., Madhura, R., and Rao, M. R. (2016). Monsoon-extratropical circulation interactions in Himalayan extreme rainfall. *Climate Dynamics*, 46(11-12):3517–3546.

Vidale, M. (2000). *The archaeology of Indus crafts: Indus craftspeople and why we study them*, volume 4. IsIAO.

Vieira, L. E. A., Solanki, S. K., Krivova, N. A., and Usoskin, I. (2011). Evolution of the solar irradiance during the Holocene. *A&A*, 531:A6.

Visbeck, M. H., Hurrell, J. W., Polvani, L., and Cullen, H. M. (2001). The North Atlantic Oscillation: Past, present, and future. *Proceedings of the National Academy of Sciences*, 98(23):12876–12877.

von Rad, U., Schaaf, M., Michels, K. H., Schulz, H., Berger, W. H., and Sirocko, F. (1999). A 5000-yr record of climate change in varved sediments from the oxygen minimum zone off Pakistan, Northeastern Arabian Sea. *Quaternary Research*, 51(1):39 – 53.

Wang, B. (2006). *The Asian Monsoon*. Springer.

Wang, J. and Wolff, D. B. (2010). Evaluation of TRMM ground-validation radar-rain errors using rain gauge measurements. *Journal of Applied Meteorology and Climatology*, 49(2):310–324.

Wang, Q., Yi, S., and Sun, W. (2017). Precipitation-driven glacier changes in the Pamir and Hindu Kush mountains. *Geophysical Research Letters*, 44(6):2817–2824.

Wanner, H., Beer, J., Bütikofer, J., Crowley, T. J., Cubasch, U., Flückiger, J., Goosse, H., Grosjean, M., Joos, F., Kaplan, J. O., Küttel, M., Müller, S. A., Prentice, I. C., Solomina, O., Stocker, T. F., Tarasov, P., Wagner, M., and Widmann, M. (2008). Mid- to Late Holocene climate change: an overview. *Quaternary Science Reviews*, 27(19):1791 – 1828.

Wasson, R., Smith, G., and Agrawal, D. (1984). Late quaternary sediments, minerals, and inferred geochemical history of Didwana Lake, Thar Desert, India. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 46(4):345 – 372.

Weber, S. (2003). Archaeobotany at harappa: indications for change. In Weber, S. and Belcher, W., editors, *Indus ethnobiology: new perspectives from the field*. Lexington Books, Lanham.

Wescoat Jr., J. L., Siddiqi, A., and Muhammad, A. (2018). Socio-hydrology of channel flows in complex river basins: Rivers, canals, and distributaries in Punjab, Pakistan. *Water Resources Research*, 54(1):464–479.

Wheeler, R. E. M. (1968). *The Indus Civilisation*. Cambridge University Press Cambridge, 3 edition.

WMO (2008). Guide to hydrological practices. volume i. hydrology–from measurement to hydrological information.

Wolff, M., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T., and Brækkan, R. (2015). Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study. *Hydrology and Earth System Sciences*, 19(2):951–967.

Woollings, T., Barriopedro, D., Methven, J., Son, S.-W., Martius, O., Harvey, B., Sillmann, J., Lupo, A. R., and Seneviratne, S. (2018). Blocking and its response to climate change. *Current Climate Change Reports*, 4(3):287–300.

Wright, R. P. (2010). *The ancient Indus: Urbanism, economy, and society*. Cambridge University Press Cambridge.

Wu, C.-H., Lee, S.-Y., Chiang, J. C. H., and Hsu, H.-H. (2016). The influence of obliquity in the early Holocene Asian summer monsoon. *Geophysical Research Letters*, 43(9):4524–4530.

Wünnemann, B., Demske, D., Tarasov, P., Kotlia, B. S., Reinhardt, C., Bloemendal, J., Diekmann, B., Hartmann, K., Krois, J., Riedel, F., and Arya, N. (2010). Hydrological evolution during the last 15kyr in the Tso Kar lake basin (Ladakh, India), derived from geomorphological, sedimentological and palynological records. *Quaternary Science Reviews*, 29(9):1138 – 1155.

Xie, P. and Arkin, P. A. (1997). Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society*, 78(11):2539–2558.

Xie, P., Chen, M., and Shi, W. (2010). CPC unified gauge-based analysis of global daily precipitation. In *Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc*, volume 2.

Xueyuan, K. and Yaocun, Z. (2005). Seasonal variation of the East Asian Subtropical Westerly Jet and its association with the heating field over East Asia. *Advances in Atmospheric Sciences*, 22(6):831–840.

Yadav, R., Ramu, D., and Dimri, A. (2013). On the relationship between ENSO patterns and winter precipitation over North and Central India. *Global and Planetary Change*, 107:50 – 58.

Yadav, R. K., Kumar, K. R., and Rajeevan, M. (2012). Characteristic features of winter precipitation and its variability over northwest India. *Journal of Earth System Science*, 121(3):611–623.

Yadav, R. K., Rupa Kumar, K., and Rajeevan, M. (2009). Increasing influence of ENSO and decreasing influence of AO/NAO in the recent decades over northwest India winter precipitation. *Journal of Geophysical Research: Atmospheres*, 114(D12).

Yadav, R. K., Yoo, J. H., Kucharski, F., and Abid, M. A. (2010). Why Is ENSO Influencing Northwest India Winter Precipitation in Recent Decades? *Journal of Climate*, 23(8):1979–1993.

Yamada, T. J., Takeuchi, D., Farukh, M. A., and Kitano, Y. (2016). Climatological characteristics of heavy rainfall in Northern Pakistan and atmospheric blocking over Western Russia. *Journal of Climate*, 29(21):7743–7754.

Yanai, M., Li, C., and Song, Z. (1992). Seasonal heating of the Tibetan Plateau and its effects on the evolution of the Asian summer monsoon. *Journal of the Meteorological Society of Japan. Ser. II*, 70(1B):319–351.

Yang, S., Lau, K.-M., Yoo, S.-H., Kinter, J. L., Miyakoda, K., and Ho, C.-H. (2004). Upstream Subtropical Signals Preceding the Asian Summer Monsoon Circulation. *Journal of Climate*, 17(21):4213–4229.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A. (2012). APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, 93(9):1401–1415.

Yu, W., Yang, Y.-C., Savitsky, A., Alford, D., Brown, C., Wescoat, J., Debowicz, D., and Robinson, S. (2013). *The Indus basin of Pakistan: The impacts of climate risks on water and agriculture*. The World Bank.

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y. (2019). Mri mri-esm2.0 model output prepared for cmip6 cmip.

Zhao, Y., Braconnot, P., Marti, O., Harrison, S., Hewitt, C., Kitoh, A., Liu, Z., Mikolajewicz, U., Otto-Bliesner, B., Weber, S., and et al. (2005). A multi-model analysis of the role of the ocean on the African and Indian monsoon during the mid-Holocene. *Climate Dynamics*, 25(7-8):777–800.

Zhu, Y. and Newell, R. E. (1998). A Proposed Algorithm for Moisture Fluxes from Atmospheric Rivers. *Monthly Weather Review*, 126(3):725–735.

Ziese, M., Rauthe-Schöch, A., Becker, A., Finger, P., Meyer-Christoffer, A., and Schneider, U. (2018). GPCC full data daily version 2018 at 1.0°: Daily land-surface precipitation from rain-gauges built on GTS-based and historic data.